

T.C.
FIRAT ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
TEKNOLOJİ VE BİLGİ YÖNETİMİ
ANABİLİM DALI



EKONOMİ HABERLERİNİN BİST 100
ENDEKSİNE ETKİSİNİN VERİ
MADENCİLİĞİ İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

DANIŞMAN
Yrd. Doç. Dr. Cem AYDEN

HAZIRLAYAN
Özlem ALPAY

ELAZIĞ-2017

T.C.
FIRAT ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
TEKNOLOJİ VE BİLGİ YÖNETİMİ ANABİLİM DALI

EKONOMİ HABERLERİNİN BİST 100
ENDEKSİNE ETKİSİNİN
VERİ MADENCİLİĞİ İLE İNCELENMESİ

YÜKSEK LİSANS TEZİ

DANIŞMAN
Yrd. Doç. Dr. Cem AYDEN

HAZIRLAYAN
Özlem ALPAY

Jürimiz 23.10.2017 tarihinde yapılan tez savunma sınavı sonunda bu yüksek lisans tezini oy birliği / oy çokluğu ile başarılı saymıştır.

Jüri Üyeleri:

1. Doç. Dr. Yusuf Cahit ÇUKRACI
2. Yrd. Doç. Dr. Özcan DEMİR
3. Yrd. Doç. Dr. Cem AYDEN



F. Ü. Sosyal Bilimler Enstitüsü Yönetim Kurulunun tarih vesayılı kararıyla bu tezin kabulü onaylanmıştır.

Prof. Dr. Ömer Osman UMAR
Sosyal Bilimler Enstitüsü Müdürü

ÖZET**Yüksek Lisans Tezi****Ekonomi Haberlerinin BİST 100 Endeksine Etkisinin
Veri Madenciliği ile İncelenmesi****Özlem ALPAY****Fırat Üniversitesi****Sosyal Bilimler Enstitüsü****Teknoloji ve Bilgi Yönetimi Anabilim Dalı****Teknoloji ve Bilgi Yönetimi Bilim Dalı****Elazığ – 2017; Sayfa: XI+75**

Bu tez çalışmasında BİST 100 borsa endeksindeki değişimler ile ulusal bazda yayımlanan haberler arasında bir ilişki olup olmadığı araştırılmıştır. Bu doğrultuda Dünya Gazetesinin resmi internet sitesi üzerinden 01.04.2016 - 01.10.2016 tarihlerinde yayınlanan ekonomi haberleri kullanılmıştır. Haber metinleri kullanılarak elde edilen veri setine metin madenciliği yöntemleri uygulanmıştır. Oluşturulan veri seti çok büyük olup üzerinde işlem yapmayı zorlaştırmasından dolayı veri seti üzerinde bazı işlemler yapılarak veri boyutu küçültülmüştür.

Öznitelik vektörleri oluşturulduktan sonra sınıflandırma işlemine tabi tutulmuştur. Sınıflandırma işlemi yapmanın temel amacı kelimelerin hangi gruba ait olduğunu belirlemektir. Kelimeler pozitif, negatif ve nötr olmak üzere 3 etikete sahiptir. Oluşturulan veri setlerine makine öğrenme algoritmaları uygulanmıştır. Bu uygulamanın sonucunda kelimelere hangi gruba ait olduklarını gösteren etiketler atanmıştır. Bu etiketlendirme sonucunda kelimeler Tf – Idf ağırlıklandırma yöntemleri uygulanarak kelimelerin ağırlık tablosu oluşturulmuştur. Bu tablodaki değer ve etiketlere bakılarak piyasanın durumu karşılaştırılmıştır.

Anahtar Kelimeler: BİST 100, Veri Madenciliği, Metin Madenciliği

ABSTRACT**Master Thesis****Investigation of the Effects of Economic News on BIST 100 Index by Using Data Mining****Özlem ALPAY****The University of Firat****The Institute of Social Science****The Technology and Knowledge Management Master Science****Elazığ-2014; Page: XI+ 75**

In this master thesis we tried to investigate whether there is a relationship between changes on BIST 100 Indeks and domestic news published on internet. In this dimension we use the news published on Dünya newspaper's web site between the 01.04.2016- 01.10.2016 dates. Data sets which obtained by using news text are reviewed with data mining. Because of huge amount of data sets, we reduced them to small size.

Feature vectors are converted to appropriate form to process in. Then these feature vectors classified in. Basic aim of classification process is to determine the belonging groups of the words. Words have three labels positive, neutral, and negative. Machine learning algorithmism have been applied to generated clusters. At the end of this application labels have been assigned to the words to determine their belonging clusters. By applying the Tf-Idf weighting methods to the words, weighting tables of the words formed. By looking at values and impacts on this table market situation compared.

Key Words: BIST 100, Data Mining, Text Mining

İÇİNDEKİLER

ÖZET	II
ABSTRACT	III
İÇİNDEKİLER	IV
ŞEKİLLER LİSTESİ	VIII
TABLolar LİSTESİ	IX
KISALTMALAR	X
ÖNSÖZ	XI
GİRİŞ	1

BİRİNCİ BÖLÜM

1. TEORİK ÇERÇEVE	2
1.1.Piyasalar	2
1.1.1. Gerçek Piyasalar.....	2
1.1.2. Mali Piyasalar.....	2
1.1.2.1.Para Piyasası	3
1.1.2.2.Sermaye Piyasası	3
1.1.2.2.1. Birincil Piyasa	4
1.1.2.2.2. İkincil Piyasa	4
1.2. Borsa.....	4
1.2.1. Menkul Kıymetler Borsaları	5
1.2.2. Menkul Kıymetler Borsalarının Tarihsel Gelişimi.....	6
1.2.3. Dünyadaki Borsalar	6
1.2.3.1. Süper Borsalar	7
1.2.3.2. Büyük Borsalar	8
1.2.3.3. Orta Büyüklükteki Borsalar	9

1.2.3.4. Küçük Borsalar	9
1.2.4. Türkiye’de Borsanın Tarihsel Gelişimi	9
1.2.5. BİST.....	11
1.3. Endeksler.....	12
1.3.1. BİST Endeksleri.....	12
1.3.1.1. Bist 30 Endeksi	12
1.3.1.2. Bist 50 Endeksi	12
1.3.1.3. Bist 100 Endeksi	12
1.4. Haberler.....	13
1.4.1. Haberlerin Piyasaya Etkisi	15
1.4.2. Borsa Tahmini ile Yapılan Çalışmalar	16
1.5. Veri Madenciliği.....	24
1.5.1. Veri Madenciliği Tanımı	24
1.5.1.1. Bilgi işlem gücü.....	25
1.5.1.2. Data tutma ve yönetme yetkinliği.....	26
1.5.1.3. Analitik araçların gelişmesi	26
1.5.2. Veri Madenciliğinin Önemi.....	26
1.5.3. Veri Madenciliği Süreci	27
1.5.4. Veri Madenciliği Kullanılan Yöntemler.....	28
1.5.4.1. Sınıflandırma	28
1.5.4.1.1. İlişki Tabanlı Sınıflandırıcılar	31
1.5.4.1.2. Karar Ağaçları.....	31
1.5.4.1.3. K-En Yakın Komşu.....	32
1.5.4.1.4. Yapay Sinir Ağları	32
1.5.4.1.5. Genetik Algoritmalar	33
1.5.4.1.6. Navies Bayes	33
1.5.4.2. Kümeleme.....	34

1.5.4.3. Tahminleyici Model.....	34
1.5.4.4. Birliktelik Kuralları	35
1.5.4.5. Ardışıklık Örüntüsü	35

İKİNCİ BÖLÜM

2. METİN MADENCİLİĞİ	36
2.1. Metin Madenciliği Çalışma Alanları.....	37
2.2. Metin Madenciliği Kullanım Alanları.....	38
2.3. Metin Madenciliği Süreci	40
2.3.1. Verilerin Toplanması.....	41
2.3.2. Veri Önışleme.....	41
2.3.2.1. Dönüşürme	41
2.3.2.1.1. Ayrıştırma Yöntemleri	42
2.3.2.2. Tarama ve İşaretleme.....	43
2.3.2.3. Kök Bulma	43
2.3.2.3.1. Cümle Çözümlemesi	43
2.3.2.3.1.1. Zemberek Kütüphanesi	43
2.3.2.4. Durak Kelimelerin Çıkarılması.....	44
2.3.3. İndeksleme.....	44
2.3.3.1. Vektör Uzay Modeli	44
2.3.3.2. Kelime Ağırlıklandırma.....	45
2.3.3.2.1. Bit Ağırlıklandırma.....	46
2.3.3.2.2. Sözcük Frekansı Ağırlıklandırma (TF)	46
2.3.3.2.3. Ters Doküman Frekansı (IDF)	47
2.3.4. Öznitelik Seçimi	47
2.3.5. Sınıflandırma	48
2.3.5.1. Kümeleme.....	48
2.4. Metin Madenciliği Aracı: WEKA	50

2.4.1. Veri Eğitimi ve Test	51
2.4.1.1. Doğruluk- Hata Oranı	52
2.4.1.2. Kesinlik	52
2.4.1.3. Duyarlılık	53
2.4.1.4. F- Ölçütü	53
ÜÇÜNCÜ BÖLÜM	
3. MATERYAL ve YÖNTEM	54
3.1. Veri Setinin Oluşturulması	55
3.2. Metin İşleme	58
3.3. Sınıflandırma	59
3.4. Kelime Ağırlıklandırma	61
SONUÇ ve ÖNERİLER	64
KAYNAKÇA	66
EKLER	74
Ek-1: Tez Orijinallik Raporu	74
ÖZGEÇMİŞ	75

ŞEKİLLER LİSTESİ

Şekil 1. Veri Madenciliğinde Kullanılan 3 Anahtar Teknolojisi.....	25
Şekil 2. Veri Madenciliğinin Etkileri	27
Şekil 3. CRISP- DM veri madenciliği süreci	28
Şekil 4. Sınıflandırma model kurma süreci	29
Şekil 5. Test Verisi Üzerinde Sınıflandırma Kurallarının Uygulanması.....	30
Şekil 6. Bir Karar Ağacının Temel Yapısı	31
Şekil 7. Metin madenciliği ile diğer disiplinler arasındaki ilişki.....	36
Şekil 8. Metin İşleme Süreci.....	40
Şekil 9. Veri ön işleme adımları	41
Şekil 10. Aynı köke sahip kelimeler.....	43
Şekil 11. Düşündü kelimesin kök ve eklerine ayrıştırılması	43
Şekil 12. Zemberek kütüphanesine göre kelime yapısı	44
Şekil 13. Vektör uzay modeli	45
Şekil 14. WEKA’da Applications Menüsü.....	50
Şekil 15. WEKA’da Veri Kümesinin Seçimi	50
Şekil 16. Model Başarı Ölçütleri	52
Şekil 17. Uygulamanın Genel Yapısı	54
Şekil 18. Haberlerin genel yapısı.....	55
Şekil 19. Dünya gazetesinin RSS uygulaması.....	56
Şekil 20. RSS ile ulaşılan haberlerin XML dosyası	56
Şekil 21. Haber dökümanının veritabanında gösterimi	57
Şekil 22. BİST 100 endeksine ait verilerin veritabanında gösterimi	58
Şekil 23. Özniteliklerin Sınıflandırılması.....	59
Şekil 24. K çaprazlama tekniğinin uygulanması	60

TABLULAR LİSTESİ

Tablo 1. Bazı Karar Ağacı Algoritmaları ve Özellikleri.....	32
Tablo 2. Metin Madenciliği Kullanım Alanları (1).....	39
Tablo 3. Metin Madenciliği Kullanım Alanları (2).....	40
Tablo 4. Örnek Metinler.....	46
Tablo 5. Sınıflandırma Algoritmalarının Karşılaştırılması	60
Tablo 6. Negatif Kelime Listesi	62
Tablo 7. Pozitif Kelime Listesi	62
Tablo 8. Nötr Kelime Listesi.....	63



KISALTMALAR

BIST	: Borsa İstanbul
CPU	: Central Processing Unit
FN	: False Negative
FP	: False Positive
GA	: Genetik Algoritmalar
GSMH	: Gayri Safi Milli Hasıla
IDF	: Inverse Document Frequency
İMKB	: İstanbul Menkul Kıymetler Borsası
KNN	: K Nearest Neighborhood
NB	: Navies Bayes
SFS	: Sequential Forward Selection
SVM	: Support Vector Machine
TCDC	: Terim-Kategoriye Bağlı Şartlar
TF	: Term Frequency
TN	: True Negative
TP	: True Positive
TS	: Term Strength
VM	: Veri Madenciliği
VTBK	: Veri Tabanlarında Bilgi Keşfi
YSA	: Yapay Sinir Ağları

ÖNSÖZ

Bu çalışma sonucunda elde edilen bilgi ve bulguların yeni çalışmalara alt yapı oluşturmasını, insanlığa bilimsel ve akılcı faydalar sağlamasını ve daha verimli çalışmaların oluşmasına katkıda bulunmasını dilerim

Tez çalışmam süresince bilgisi ve vaktini benden esirgemeyerek bana yol gösteren değerli danışmanım ve hocam Yrd. Doç. Cem AYDEN'e sonsuz teşekkürlerimi sunuyorum.

Özlem ALPAY

ELAZIĞ - 2017



GİRİŞ

Bilgisayarların kullanımının artması ile beraber bilgi ve bilginin işlenmesi önem kazanmıştır. Bilgi ve teknolojinin birlikte kullanılmasını ifade eden bilişim, bilgi ve teknoloji arasında köprü kurarak hayatı kolaylaştırmaktadır. Yoğun rekabetin yaşandığı ve bilginin öneminin her geçen gün arttığı günümüzde, doğru ve geçerli bilgiyi elde eden organizasyonlar büyük rekabet avantajı elde ederler. Bu rekabet ortamındaki başarı ölçütü de genellikle sahip olduğu ekonominin durumu olarak ifade edilir.

Günümüz dünyasında herhangi bir devletin sahip olduğu tüm güçler neredeyse ekonomisi ile doğru orantıdadır. Bir devletin ekonomisi ne kadar güçlü ise o ülke dünya da o kadar söz sahibidir. Bu gibi durumlar gittikçe arttığından insanları ekonomiye bunu etkileyen faktörlerin neler olduğu konusunda merak ettirmektedir. Ülkelerin ekonomilerini etkileyen gerek küresel gerekse yerel nedenler bulunmaktadır. Bu yerel nedenler üzerinde de ülke içinde gerek yaşanan siyasi, kültürel gerekse ekonomik gelişmelerin etkileri vardır. Bu etkileri inceleyip anlayabilme bilişim dünyasının ilgisini çeken ve üzerinde çalışmaların yapıldığı konulardan biridir. Bu konunun bu kadar ilgi görmesinin nedenlerinden biri de yatırımcı algısıdır. Yatırımcıların hedefleri arasında kazanma algısı vardır bu da yatırımcıyı piyasaya hakkında doğru ve başarılı tahminler bulmaya yöneltmiştir.

Bu konu üzerinde bilişim yöntemlerinden yapay sinir ağları, veri madenciliği vb. yöntemleri kullanarak birçok çalışma yapılmıştır. Bu tez ile BİST 100 endeksine ekonomi haberlerinin etkisini inceleyen bir çalışma yapılmıştır. Ekonomi haberleri kullanarak oluşturulan öznitelikler haberler içerisinde geçen kelimelerdir. Bu kelimelerin BİST 100 endeksinin üzerindeki etkileri incelenecektir.

BİRİNCİ BÖLÜM

1. TEORİK ÇERÇEVE

Bu bölümde Borsa, BİST 100 Endeksi, veri madenciliği, haberler ve haberlerin sınıflandırılması gibi çerçeveyi oluşturan teorik konulara değinilmiştir.

1.1. Piyasalar

Piyasa, alıcı ve satıcıların bir araya gelerek iletişim kurması ile belirli alım ve satım işlemlerini gerçekleştirdikleri ortamdır (http://en.karsan.com.tr/images/yatirimci_iliskileri/s_sorulan_sorular/sorularlar_borsa_sermaye_piyasasi.pdf).

Piyasalar; iktisadi açıdan gerçek piyasalar ve mali piyasalar olmak üzere ikiye ayrılır.

1.1.1. Gerçek Piyasalar

Üretilmiş mal ve hizmetler ile bu mal ve hizmetlerin üretiminde kullanılan üretim faktörlerinin karşılaştıkları piyasalardır. Otomobil piyasası, pamuk piyasası vb. piyasalar bu kapsamdadır (http://en.karsan.com.tr/images/yatirimci_iliskileri/s_sorulan_sorular/sorularlar_borsa_sermaye_piyasasi.pdf).

1.1.2. Mali Piyasalar

Mali piyasalar; “mali varlık” olarak tanımlanan değerli evrak niteliğindeki belgelerin alınıp satıldığı piyasalardır. Mali (Finansal) piyasa kavramı, fon arz edenler, fon isteyenler, fon akımlarını tertip eden müesseseler, akımı sağlayan araç ve gereçler ile bunları tertip eden adli ve yönetsel kaideleri içerir. Finansal piyasa; para ve sermaye piyasalarından daha geniş ve bu piyasaları da içine alan bir kavramdır.

Finansal piyasalar aşağıdaki unsurlardan oluşmaktadır.

- a) Tasarruf sahipleri (fon arz edenler),
- b) Tüketiciler/Yatırımcılar (fon talep edenler),
- c) Yatırım ve finansman araçları,
- d) Yardımcı kuruluşlar,

e) Hukuki ve idari düzen

Finansal piyasalar piyasaların yerine göre organize ve organize olmayan piyasalar olarak ayrılırlar.

Türkiye’de organize piyasalar T.C. Merkez Bankası Piyasaları (Açık Piyasa İşlemleri, Para Piyasası, Döviz Piyasası), Borsa İstanbul Piyasaları, Takasbank Piyasaları, Bankalararası TL Piyasası, Bankalararası Repo Piyasası gibi.

Türkiye’de organize olmayan piyasalar olarak ise Bankalararası Tahvil Piyasası, Bankalararası Döviz Piyasası, Serbest Döviz Piyasası ve Serbest Altın Piyasası sayılabilir. Mali piyasalar da Para Piyasası ve Sermaye piyasası olarak ikiye ayrılmaktadır.

1.1.2.1. Para Piyasası

Para piyasası, vadesi en fazla bir yıl olan arz, talep ve fonların karşılandığı piyasalardır. Vadesi bir yıldan az olması ile söylenmek istenen varlığın likit olmasıdır. Para piyasasında vade sürelerinin az olmasından dolayı bu piyasa türünde alınıp satılan borç araçlarının arz fiyatlarında çok fazla dalgalanmalar yaşanmaz. Bundan dolayı para piyasası araçlarının riski azdır.

Bu araçlar aşağıdaki gibi listelenebilir:

- Hazine Tahvilleri
- Mevduat Sertifikası
- Finansman Bonosu
- Repo
- Çek ve Senetler

1.1.2.2. Sermaye Piyasası

Sermaye piyasası, kısa orta ve uzun vadede menkul değerlere fonlara arz ve taleplere bağlı olarak gelişen piyasalardır. Bu piyasalarda da fonlar, menkul değerler ticaret yoluyla el değiştirmektedirler.

Sermaye piyasası bu işlevini, kendi kuralları içinde bazı araçlar kullanarak yürütür. Bu araçlar, menkul değerler ve diğer sermaye piyasası araçlarıdır (Aykut, 2015: 6).

Menkul Değerler: Tahvil ve hisse senetleri çeşitleri, kar ve zarar ortaklık belgeleri, yatırım fonları katılım belgeleri, hazine, banka ve finansman bonoları, yeni pay alımı ve tahvil faiz kuponları, katılma intifa ve gelir ortaklığı senetleri, gayrimenkul sertifikaları gibi sıralanabilir .

Diğer Sermaye Piyasası Araçları: Menkul değerlerin dışında kalan şartları kurulca belirlenen değerli evraklardır. Bu araçlara örnek, faiz ve temettü kuponları, opsiyon senetleri, mevduat sertifikaları verilir.

Sermaye piyasası birincil ve ikincil piyasa olmak üzere ikiye ayrılır.

1.1.2.2.1. Birincil Piyasa

Pay senedi, menkul değerleri ve borçlanma araçlarını satıcı şirketler ile alıcıların doğrudan muhatap oldukları ve şirketlerin gerektiğinde harcanabilecek finansal kaynakları yarattıkları piyasalardır. Bu piyasalarda yatırımcıya arz kuralları uygulanır.

1.1.2.2.2. İkincil Piyasa

Bu piyasalarda, finansal araçların menkul değer miktarını artırmak suretiyle, arzını kolaylaştırır ve nakit sıkıntısının giderilmesine yardımcı olur. Menkul kıymet borsaları, ikinci el senetlerin işlem gördüğü en önemli ikincil piyasalardır (Aykut, 2015: 11).

1.2. Borsa

Borsa, kendine özel belirli kuralları ve yasalara uygun kriterlere sahip, arz-talep ilişkisine göre fiyatları belirlenen farklı ürün ve varlıkların ticari faaliyetlerinin yapıldığı piyasalardır.

Borsanın iç yapısında bulunan bu kıymetli varlıklara menkul değerler denir. Menkul değerler; kıymetli evraklar, ticari mallar, işlenmemiş ürünler, hisse senetleri ve yabancı paralar olarak çeşitlilik gösterir. Bu varlıkların ticaretlerinin yapılması ve arz-talep dengesi sonucu ortaya çıkan doneler borsanın taban bilgilerini oluşturur (Özyurt, 2015: 1).

Borsaların kuruluş amaçları, sermaye piyasası araçlarına yönelik alımları, kısa vadeli fon imkanlarını ve güvenli fiyat oluşumunu sağlayarak açık, etkili, güvenilir, erişilebilir bir piyasa oluşturmak, ve bunu geliştirmektir.

Borsalar, piyasa araçlarını yatırımcı ve aracı kurumlar ile diğer aktörlerin

hizmetine sunmak suretiyle ülke ekonomisinin gelişmesi ve yatırımlar için gerekli olan finansman desteği sağlamak amacıyla kurulan organize piyasalardır (Aykut, 2015: 12).

Borsalar bazı özelliklere sahiptirler, şöyle ki,

- Değiş- tokuş edilebilir,
- Şeffaflık ön plandadır,
- Likidite sağlar,
- Arz ve taleplere bağlı olarak fiyatlar oluşur. Oluşan bu fiyatlar; bütün piyasalarda geçerlidir,
- Sektör ve şirketler ekonomik durum hakkında bilgilendirme yaparlar,
- Finansal destek isteyenlere kısa zamanda sağlıklı ortam sunmak suretiyle yatırımcıya karlı yatırım fırsatları yaratırlar,
- Ülkede çalışan sayısının fazlaşmasına katkı sağlamak gibi özellikleri sayabiliriz.

Borsalarda, şirketlerin her türlü hareketlerinin izlenmesinin yanında piyasaları izleyebilmek amacıyla çeşitli şekillerde endeksler hesaplanmaktadır. Bu amaçla BİST piyasasına ait toplam 324 endeks hesaplanmaktadır. (54 anlık endeks, 270 seans sonu endeks)

Hisse senedi endeksleri, iki şekilde hesaplanmaktadır.

- a) Getiri endeksi, kar paylarının ödemelerini esas alır.
- b) Fiyat endeksi, fiyatlardaki değişimi esas alır.

Bunların dışında diğer piyasalara dönük de endeksler hesaplanmaktadır. Bu araçların (altın, fon, borçlanma, mevduat) getirilerini hesaplayabilmek için BİST – KYD adıyla 29 endeks hesaplanmaktadır (Telli, 2016: 3).

1.2.1. Menkul Kıymetler Borsaları

Sermaye piyasalarının ana unsuru olan menkul kıymetler borsası belirlenmiş kurallara göre kayıt altına alınmış menkul kıymetlerin alım ve satımlarının yapıldığı yerlerdir. Bu alım ve satım işlemleri yasal kurallar çerçevesinde yapılır.

Borsada, ticari faaliyetler direkt bu kıymetleri elinde tutanlar tarafından yapılır. Bu elemanlar farklı isimlerle (Banker, acente, broker, jobber vb.) farklı statüde ya aracı olarak ya da kendi adına görev yaparlar.

Borsadaki işlemler, piyasa araçlarının alım ve satımı için görevlendirilen aracıya borsa yetkilerinin verilmesiyle başlar.

Menkul Kıymetler Borsası, sermaye fazlası olan kurumlardan sermaye açığı olan

kurumlara aktarılacak fonların alınıp satıldığı, alıcı ve satıcıları bir araya getiren borsalara denir (Demirel, 2009: 25).

1.2.2. Menkul Kıymetler Borsalarının Tarihsel Gelişimi

Menkul kıymetler borsasının ortaya çıkmasına neden olan etkenlerin başında 15 yy.' da denizcilikteki gelişmeler sonucu coğrafi keşiflerin etkisiyle ticari faaliyetlerin gelişmesi ve vadeli ürün satışlarının artmasıdır.

Belçikalı tüccarlar ürün satışı sonrasında firmadan aldıkları vadeli ticari senetleri vadesinden önce peşin paraya çevirmek için para ticareti yapan kurum veya kişilere bu ticari senetler üzerindeki rakamlardan indirim yaparak senetlerin alım satımını gerçekleştirmişlerdir. İndirimli senedi alan kurum ve kuruluşlar da tasarruf etmek isteyen kişilere bu senetleri satması sonucu menkul kıymetler el değiştirmiş olur. Bu ticari faaliyetlerin zamanla gelişmesi bu alanda güvenli bir mekanın kurulması düşüncesini ortaya çıkarmıştır.

Bu çalışmalar sonucunda değerli ürünlerin alınıp satılması disiplinli ve yasal yükümlülük garantisi olan menkul kıymetler borsası Dünya da ilk olarak Belçika'nın Anvers kentinde kurulmuş oldu.16. yy'da Amsterdam Borsası (Hollanda) Anvers borsasının yerini almıştır. 16. yy'dan itibaren Avrupa'nın diğer bölgelerinde de yeni borsalar kurulmaya başlamıştır.

Menkul kıymet borsaları, günümüzde birçok ülkede mahalli, ulusal ve uluslararası alanda çalışmalarını yürütmektedirler.

- Mahalli borsalar, buldukları bölgedeki yerel firmaların piyasa araçlarının ticaretine piyasa oluşturur.
- Ulusal borsalarda, ülke içinde faaliyetler yapılır.
- Uluslararası borsalar, ülke dışında ki borsalarda yapılan faaliyetlerdir. Önemli uluslararası borsalar, New York, Londra, Tokyo, Paris, Toronto borsalarıdır.

Yatırımcılar, aracı kurum vasıtasıyla yerel, ulusal ve uluslararası borsalardan alım ve satım gerçekleştirebilirler.

1.2.3. Dünyadaki Borsalar

Borsaların gelişmişlik düzeyleri buldukları ülkelerin ekonomik, siyasi ve sosyal yapılarına göre şekillenmektedir. Borsalar büyüklükleri göz önüne alınarak sınıflandırılmaktadır. İşlem ve piyasa değeri özelliklerine göre borsalar;

- Süper
- Büyük
- Orta
- Küçük borsalar olarak adlandırılabilir.

Borsaların bir diğer sınıflandırma şekli ise, Dünya Bankasının verilerine göre borsalar, Dünya Bankasının ülkelerin ekonomik yapısını dikkate alarak yaptığı sınıflandırmadır. Bu sınıflandırmada o ülkenin kişi başı Gayri safi milli hasıla (GSMH) göz önüne alınarak yapılan sınıflandırmadır. GSMH baz alınarak yapılan sınıflandırmada, 12.276 ABD doları esas alınmış.

- Ülkede kişi başı GSMH'sı 12.276 ABD doların üstünde ise gelişmiş borsa,
- Ülkede kişi başı GSMH'sı 12.276 ABD doların altında ise gelişmekte veya gelişen borsa olarak adlandırılır.

Gelişmiş borsalar genellikle sanayide ileri ülkelerdeki büyük borsalardır. Dünyanın en büyük borsası NYSE (NewYork Stock Exchange), ikincisi Nasdaq borsasıdır.

Ülkemizdeki BİST gelişmekte olan borsalar içerisinde yer almaktadır. Bu gruba dâhil borsalar genellikle gelişmekte olan ülkelerde bulunurlar. Bu borsaların önem arzedenleri Brezilya, Arjantin, Endonezya, Malezya, Türkiye BİST, Rusya gibi gelişmekte olan ülke borsalarını sayabiliriz.

1.2.3.1. Süper Borsalar

New York Borsası (NewYork Stock Exchange) NYSE: Bu borsa gönüllü bir grup insanın kurduğu dernekle ortaya çıkmış. Bu borsanın yönetim kurulu borsanın kendi üyeleri tarafından seçilir. Bu borsanın üyeleri şunlardır;

Komisyoncular (brokers): Yatırımcının işlem isteğini specialistlere iletir ve karşılığında komisyon alır.

Uzmanlar (Specialistler): Borsa konusunda bilgi, görüş ve becerisi olan specialistler müşteri isteklerini getiren komisyoncuların brokerliğini yapan borsa üyeleri

Tacirler(dealers): Kendi adlarına borsada alım satım yapan borsa üyeleri.

ABD en fazla işlem gören borsa NYSE'dir. Ülkedeki borsa işlem hacminin % 85, i bu borsada işlem görmektedir. NYSE de menkul değerlerini kote etmek (kayıt) isteyen şirketlerin NYSE nin bazı şartlarını taşımaları gerekir.

Tokyo Borsası (Tokyo Stock Exchange) TSE: Uzak Doğunun en önemli borsasıdır. Borsa üyeleri düzenli ve saitori'ler olarak gruplara ayrılmıştır.

Düzenli üyeler: Kendi ve müşteri adına alım satım yapan borsa üyeleri

Saitori'ler: Düzenli üyeler arasındaki işlemlerde uzman aracı olarak görev yaparlar.

NYSE Borsası'nın specialistlerine benzetilmektedirler. Tokyo Borsası teknolojik alt yapısıyla önemli bir borsa olup bilgisayar ile borsa işlemleri desteklenmektedir.

Tokyo Borsası şirket olarak çalışmalarını yürütür. Borsada kamunun sıkı denetimi vardır. Borsa ile alakalı kararları maliye bakanlığı almaktadır.

Menkul Kıymetler Kurulu: Maliye bakanlığına bağlı olarak çalışan bu kurul doğrudan borsayı ve aracı kurumları denetleyebilmektedir.

Londra Borsası (London Stock Exchange) LSE: Londra'da 1801 yılında kurulmuştur. Borsanın gelişmesi için birçok değişiklik ve düzenlemeler yapılmıştır. Bunlar içinde en önemli düzenleme 1986 yılında yapılan Büyük Patlama (Big Bang) dir. Bu düzenleme ile Londra (LSE) Borsasında büyük oranda değişiklikler yapılmıştır. LSE' nin 4 pazarı bulunmaktadır;

Londra Opsiyon Pazarı,

Birleşik Krallık Hisse Senetleri,

Uluslararası Hisse Senetleri ve Kamu Menkul Kıymetleri,

Sabit faizli menkul kıymetler.

1.2.3.2. Büyük Borsalar

Amerikan Borsası (American Stock Exchange) AMEX: Amerikan Borsası'nda küçük ve yeni kurulan işletmelerin menkul kıymetleri alınıp satılır. ABD'nin ikinci büyük borsasıdır. Bu borsada alım ve satım emri düzenli ve yardımcı üyelere verilir. Bankalar borsa üyesi olmamasına rağmen bu emirleri kabul edebilirler. Bankalar borsaya ve üye şirketlere bu emirleri ulaştırarak yatırımcıya destek olmaktadır.

- Chicago Borsası (Midwest Stock Exchange)
- Pasifik Borsası
- Hong Kong Borsası,
- Paris Borsası,
- Milano Borsası,

- Frankfurt Borsası

-

1.2.3.3. Orta Büyüklükteki Borsalar

ABD: ABD'de 14 borsa bulunmaktadır(7 aktif). Boston, Cincinati, Philadelphia ve Chicago'daki borsalar.

Avrupa:

- ALMANYA, Frankfurt, Berlin, Hamburg, Münih,
- İSPANYA, Madrid, Barcelona,
- BELÇİKA, Brüksel,
- PORTEKİZ, Lizbon,
- DANİMARKA, Kopenhag,
- İSVEÇ, Stockholm,
- HOLLANDA, Amsterdam.

Uzak Doğu Borsaları: Seul, Taiwan, Singapur, Sydney, Melbourne, Bombay ve Hindistan'da bulunan toplam 14 borsa vardır.

Latin Amerika Borsaları: Rio de Janeiro, Buones Aires, Sao, Mexico City.

Kanada Borsaları: Toronto.

Orta Doğu Borsaları: İstanbul

1.2.3.4. Küçük Borsalar

Dünya'da 65'e yakın küçük borsa bulunmaktadır. Atina, Viyana, Helsinki, Kuveyt, Oslo, Nijerya (<http://www.borsahocasi.net/2015/02/dunyaborsalari.html>).

1.2.4. Türkiye'de Borsanın Tarihsel Gelişimi

Türkiye'de borsanın gelişimini cumhuriyet öncesi ve cumhuriyet sonrası olarak ikiye ayırabiliriz.

Osmanlı İmparatorluğunda Borsa faaliyetlerine Tanzimat'tan sonra rastlamaktayız. Bu dönemde Avrupa ülkeleriyle serbest ticaret anlaşmaları imzalanmış ve bu anlaşmalar ile Osmanlı İmparatorluğuna ait topraklarda malların alım- satım hakkı diğer ülkelere verilmiş oldu. Bu dönemde ilk defa İngilizlere, sonrada diğer Avrupa ülkelerine bu haklar tanınmıştır. Bu anlaşmalar sonucunda Osmanlı İmparatorluğu mali açıdan büyük oranda gelir kaybı yaşamıştır. Bu dönemde yaşanan

1854 yılındaki Kırım Savaşı Osmanlı İmparatorluğunun bütçesinide olumsuz etkilemiş ve bu savaşın sonucunda ilk defa Osmanlı İmparatorluğu dış borç almak zorunda kalmıştır.

Osmanlı İmparatorluğu mevcut giderlerin ve yatırımların finansal kaynaklarını sağlayabilmek için iç borçlanma yapmış ve tahvil çıkarmıştır. Osmanlı İmparatorluğu'nun ihraç ettiği tahviller sonucunda İstanbul'da ikincil piyasa oluşmuş oldu. Bu sermaye piyasasının işlemleri genelde Galata semtinde bulunan gayrimüslim bankerler tarafından yapılmıştır. 1866 yılında gayri müslim bankerlerin ve Osmanlı İmparatorluğunun borçlu olduğu devletlerin yardımıyla oluşturulan ve Fransız borsa sistemine yakın olan Der- Saadet Tahvilat Borsası ilk Osmanlı Borsası olarak kurulmuş oldu.

Osmanlı İmparatorluğu borsasındaki yenilikçi çalışmalar devam ederek 1904 yılında yeni borsa kuralları hazırlanmış ve 1906 yılında borsanın adı Esham ve Tahsilat Borsası ismini almıştır. Bu borsanın ömrü fazla olmadı, çünkü I. Dünya Savaşı sonrası Osmanlı İmparatorluğunun durumu Esham ve Tahvilat Borsasının kapanmasına yol açmıştır (Şahin, 2014: 15).

1923 te Cumhuriyetin ilanıyla birlikte yeni Türkiye Cumhuriyeti devletinde borsa ile ilgili çalışmalar yapılmış olup, 1929 yılında 1447 sayılı kanunla "İstanbul Menkul Kıymetler ve Kambiyo Borsası" adıyla yeni bir borsa kuruldu. Bu borsada yöntemler daha çok Fransız borsa yöntemlerine yakın uygulamalar yapılmış ve Maliye Bakanlığının görevlendirdiği belirli sayıdaki kuruluşlar eliyle yürütülmüştür. Bir süre sonra borsada İngiliz-Fransız karışımı yöntemler uygulanmaya başlanmıştır.

Türk parasını koruma kanununun 1939 yılında kabul edilmesiyle getirilen Kambiyo denetimi daha çok yabancı şirketlerin finansal varlıklarında ve Avrupa borsalarıyla birlikte çalışan, konvertibl edilmiş Türk parasına dayalı işlem gören İstanbul borsası için büyük bir olumsuzluk yaratmıştır. İstanbul borsası üzerinde ikinci bir şok da 1938 yılında borsa merkezinin Ankara'ya taşınması olmuştur. Zaman içerisinde Ankara'da ki ticari ve sanayi faaliyetlerin yetersiz finansal aktivitelerin zayıf kalacağı anlaşılmıştır. 1941 yılında yeniden borsa merkezinin İstanbul'a taşınması kararlaştırılmıştır.1981 de Sermaye Piyasası kanununun çıkması sonucu borsanın önem kazanması için bazı tedbirler alınması ihtiyacı hissedilmiştir.

Bu tedbirlerin başında 1929 yılında çıkarılan 1447 sayılı kanunun yerini almak üzere 6 Ekim 1983 yılında Kabul edilen 91 sayılı Menkul Kıymetler Borsaları, Kanun

Hükmünde Karaname ve Menkul Kıymetler Borsalarının Kuruluş ve Çalışmaları Hakkında Yönetmeliktir. 31 Ekim 1985 tarihinde İstanbul Menkul Kıymetler Borsası yönetmeliği kabul edilerek borsanın 26 Aralık 1985 tarihinde açılması sağlanmış ve 2 Ocak 1986 da ise ilk seansını başlatmıştır (<http://borsanasiloyunanir.co/borsa-tarihi/>).

1.2.5. BİST

6 Ekim 1983'te Resmi Gazetede 18183 sayısı ile yayımlanan "Menkul Kıymetler Borsaları Hakkında 91 sayılı KHK", Türkiye de bulunan sermaye piyasalarını daha aktif hale getirebilme perspektifi ile Türkiye'deki menkul kıymetlerin alım ve satım gibi işlemlerinin gerçekleştirilebilmesine uygun ortam oluşturmak amacıyla Türkiye'deki borsaların kuruluş şart ve esaslarını düzenleme görevine sahip olan ve ilk İMKB adıyla 6 Ekim 1984 tarihinde ve 18637 sayılı Resmi Gazetesinde yayımlanan "Menkul Kıymetler Borsalarının Kuruluş ve Çalışma Esasları Hakkında Yönetmelik" ile Türkiye'de bulunan menkul kıymetler borsalarının kuruluş şartlarını, çalışma şeklini, denetlenmesini, borsada alım satım esaslarını ve borsa üyelerinin kuruluş şartlarını, çalışma şekilleri ile yükümlülükleri düzenlenmiştir.

5 Nisan 2013 tarihinde İstanbul Menkul Kıymetler Borsası (İMKB) ismi "Borsa İstanbul" olarak değiştirilmiştir. Slogan olarakta "Yatırıma değer!" resmi slogan olarak kabul edilmiştir.

Borsa İstanbul, genel kurul tarafından seçilmiş yedi üyeden oluşan bir yönetim kurulu ile yönetilir.

BİST, çalışma şekil ve esasları, 19 Şubat 1996 tarihli ve 22559 sayılı Resmî Gazete'de yayımlanan "İstanbul Borsası Yönetmeliği" ile düzenlenmiştir.

Bir meslek kuruluşu olan BİST, yatırım ve kalkınma bankaları, ticari bankalar ve aracı yapılardan oluşmuş üyelere sahip olan ve hisse senetleri piyasasında sürekli olan 4 pazarı bulunan bir yapıdır. Bunlar ulusal pazar, ikinci ulusal pazar, yeni ekonomi pazarı ve gözaltı pazarıdır. BİST'in tahvil ve bono piyasasında da kesin alım - satım pazarı, repo-ters repo pazarı ve gayrimenkul sertifikaları pazarı vardır.

Borsada ortaklık hakkı ile alacaklılık hakkı tanıyan ve sermaye piyasası kurulunca sermaye piyasası araçları işlem görebilmektedir. BİST'de hisse senetleri, devlet tahvili ve hazine bonoları da işlem görmektedir.

24 Şubat 2017 tarihli yayımlanan kararname de devlete ait bütün hisselerinin Türkiye Varlık Fonuna devredilmesi kararı verilmiştir

(https://tr.wikipedia.org/wiki/Borsa_Istanbul).

1.3. Endeksler

Endeksler, belirli bir zaman aralığında fiyat, maliyet, satış başarısı gibi verileri elde etmek için oluşturulmuş göstergelerdir. Hazır Giyim Tüketim Endeksi, Sanayi Üretim Endeksi örnek olarak verilebilir (<http://www.endeksnedir.com/>).

1.3.1. BİST Endeksleri

Bist endeks türleri açıklanmış ve bu tez çalışmasında BİST 100 endeksi kullanılmıştır.

1.3.1.1. Bist 30 Endeksi

Ulusal pazarlarda işlem gören şirketler, kurumsal ürün pazarında işlem gören taşınmazların yatırım ortaklıkları ile girişim sermayesi yatırım ortaklıkları arasından seçilmiş olan 30 paydan oluşur (<http://bankamufettisleri.blogspot.com.tr/2013/07/bist-ve-bist-endeksleri-nedir.html>).

1.3.1.2. Bist 50 Endeksi

Ulusal pazarlarda işlem gören şirketler, kurumsal ürünler pazarında işlem gören taşınmazların yatırım ortaklıkları ile girişim sermayesi yatırım ortaklıklarının hisse senetlerindeki, belirli şartlara göre seçilmiş 50 paydan oluşmaktadır. BİST 30 Endeksi'nde bulunan hisse senetlerini de direk kapsamaktadır (<http://www.bireyselemeklilikbes.com/bist-50-endeksi-nedir/>).

1.3.1.3. Bist 100 Endeksi

BİST 30 ve BİST 50 endekslerinde bulunan hisseleri ve bunların yanında yine ulusal pazarda ve kurumsal ürünler pazarında işlem gören taşınmaz yatırım ortaklıkları ile girişim sermayesi yatırım ortaklıklarının hisse senetlerinden seçilmiş yüz şirket hisse payından oluşmaktadır.

Bist 100 endeksi; Borsa İstanbul Piyasasında temel endeks olarak kullanılır ve Borsa İstanbul değeri denildiğinde esas alınır (Aykut, 2015: 18).

Borsa İstanbul'da işlem gören hisselerin, senede 4 kere gözden geçirilmesi ile,

BİST 100 Endeksinde yer alacaklar belirlenir. Bu seçim, ulusal pazarda işlem gören şirketler, kurumsal ürünler pazarında işlem gören taşınmaz yatırım ortaklıkları ile girişim sermayesi yatırım ortaklıklarının payları arasından yapılır. Paylar aktif dolaşımında olan payların piyasa değerlerine ve gün içindeki ortalama işlem yoğunluğuna bağlı olarak büyükten küçüğe doğru sıralanır. Sıralamaların her ikisinde de üst sıralardaki paylar endekse alınır. Dönem içindeki değişiklikler, ilgili endeks döneminin başından yaklaşık 10 gün önce ilan edilmelidir. Ayrıca BIST 100 Endeksi, ağırlık sınırlamasız ile son işlem fiyatları ve ağırlık sınırlamalı ile ağırlıklı ortalama fiyatlar kullanılmasıyla da hesaplanmaktadır(www.borsaneder.tv/piyasalar/pay-piyasasi/pay-piyasasi-endeksleri/bist-100-endeksi/).

Bist 100 endeksi sabit türde bir endeks olmamakla beraber zaman zaman yeniden hesaplanması ilk 100 şirket değişmektedir. 1996 yılındaki değişiklik ile 02.01.1997den itibaren yeni usullere göre hesaplanmış yeni endeksin başlama değeri 988,93'dür (Şahin, 2014: 16).

1.4. Haberler

Günümüze kadar, haber üzerine pek çok tanımlar yapılmıştır. Bunlardan bazılarını şöyle sıralayabiliriz:

“Acele kaleme alınmış edebiyattır. “

“Bir olay hakkında alınan veya verilen bilgidir. “

“Bir olayın raporudur”

“Bir olayın, objektif bir biçimde anlamlandırılmasıdır. “

“Dün bilmediğimiz her şeydir. “

“Gazeteye basılıp okuyucuya aktarıldığında yararlı olacağı düşünülen olaydır. “

“Halkın ilgisini çekebilen her şeydir. “

“İnsanların bilgi edinme gereksinimini karşılayan bir çeşit öğedir. “

“İnsanların bilmek isteyecekleri her şey haberdır. “

“Tarihin kabaca yazılmış olan ilk müsveddesidir. “

“Toplumu ilgilendiren olay, düşünce ve kanıların; tam, eksiksiz ve doğru bir şekilde yayın araçları kullanarak topluma aktarılmasıdır. “

Yapılan bu tanımlardan sonra haberin akademik olarak tanımını yapacak olursak, “Haber, zamanında verilen, toplumun çoğunu ilgilendiren ve etkileyen, bu toplumdaki kişilerin anlayabileceği bir olay ya da kanaattir.” Bu tanım 1925 tarihinde

Washington Üniversitesi dekanlarından biri olan M. Lyle Spencer tarafından yapılmış en kalıcı ve genel kabul görmüş tanımdır.

Haberler konu başlıklarına göre aşağıdaki gibi sınıflandırılabilirler;

- Politika,
- Dünya olayları,
- Güncel olaylar,
- Ekonomi,
- Diğer (Magazin, Spor, Yaşam, Kültür-Sanat ...).

Politika Haberleri: Politika haberleri içinde, ulusal ve uluslararası siyasi aktörlerin söylemleri, tutumları, kararları ve davranışları yer alır. 2003 yılında yapılmış olan bir çalışmada, Amihud ve Wohl, Saddam Hüseyin'in devrilmesi ile piyasalardaki beklentiler arasındaki ilişkiyi incelemişlerdir. Bu araştırmanın sonuçlarında; Saddam Hüseyin'in devrilme ihtimalinin artması ile olumlu bir hava oluşmuş, hisse senetlerinin arttığı, Dolar'ın Euro karşısında değer kazandığı ve petrol fiyatlarının indiği görülmüştür (Amihud & Wohl, 2003).

Dünya Haberleri: Dış kaynaklı gelişmelerin olduğu bu grupta, diğer ülkelerin ekonomik faaliyetleri ile diğer faaliyetleri (uluslararası ilişkiler, askeri, enerji, yatırım, teknoloji vb.) yer alır.

Uluslararası şirket birleştirme haberlerinin, borsaya, olumlu etkileri olacaktır. ABD'nin kısa bir süre önce Parasal Genişleme Programı'nı durdurması, (ilkbaşlarda tahvil alımlarını azaltması) altın ve gelişmekte olan piyasalara olumsuz bir etki yaratmış ve Dolar'ın yerel para birimlerine karşı değer kazandığı görülmüştür.

Güncel Haberler: Genellikle ekonomik bir değere sahip olmayan, yaşamdan örnekler içeren ve çoğunlukla Steiler'in de anlattığı gibi olumsuzluk içeren haber grubudur. Gösteriler, eylemler, kazalar, istatistiki bilgiler, eğitim vb. haberler bu grupta yer alır.

Ekonomi Haberleri: Küresel piyasalar ve piyasa beklentileri, yatırımlar, ticari faaliyetler, ülkelerin finansal durumları, ithalat ve ihracat gibi finansal yoğunluklu haberlerin yer aldığı gruptur.

Diğer Haberler (Magazin, Spor, Yaşam, Kültür-Sanat): Ekonomik bir değere sahip olmayan, genellikle yaşamdan örnekler sunan haber grubudur. Bu grupta bulunan

spor haberleri istisna bir durumdur. Genellikle şampiyonluklar, uluslararası başarı ve transferler pozitif etki yaparken, başarısızlıklar negatif yönde bir etki oluşturmaktadır.

1.4.1. Haberlerin Piyasaya Etkisi

Borsa hareketleri incelendiğinde, yapılan açıklamalar, siyasi, politik, ekonomik gelişmeler ve bunlardan dolayı yatırımcıların psikolojik tepkilerinin borsanın dalgalanmalarına sebep olduğu görülmektedir. Daha ayrıntılı bir şekilde inceleme yapıldığında insan psikolojisinin borsa üzerinde büyük etkilere sahip olduğu görülecektir. Yatırımcıların var olması ile borsanın da var olacağı düşünülecek olursa, psikolojinin borsa üzerinde en büyük etkiye sahip olduğu daha iyi anlaşılacaktır.

Borsa gündem olaylarından, yatırımcıların verdikleri tepkilere göre etkilenecektir. Örneğin ABD’de bir veri açıklanmadan önce durgun bir seyir izleyen, piyasaların yatırımcıların açıklamalarına göre yön arayışına girmesidir. Yatırımcıların pozisyonlardan uzak kalması, borsanın yatay seyirde ilerlemesine neden olacaktır. Buna en iyi örnek olarak FED’in varlık alımlarını bitirmesi ve bunun sonunda yaşananları gösterebiliriz. Varlık alımlarının sonlandırılması, Amerikan ekonomisinin iyi bir durumda olduğunu gösterir ve bu da doların değerlenmesine sebep olur. Doların aşırı değer kazanması da uluslararası piyasalarda nasıl bir etkiye sahip olacağını gözlenmesi ile de düşüslere neden oluyor.

Borsa hareketlerinin hızlanma veya durgunlaşmasında etkili olan gündem olaylarında jeopolitik risklerin de büyük bir etkiye sahip olacağını da unutmamak gerekir. Genel de Çin, ABD, Japonya gibi güçlü ekonomiler incelendiğinde, dünya genelinde yaşanan jeopolitik risklerin borsayı etkilediği görülmektedir. Ülke ekonomilerinin gücü bilindiği üzere borsaların daha sağlam bir temelde olmasına olanak sağlıyor. Bu sağlam temel de borsaların gündemde yer alan olaylar karşısında nasıl etkileneceğini bizlere gösteriyor.

Türkiye piyasalarında Darbe girişimi 2016 yılının en önemli gündem maddesi oldu diyebiliriz. Yatırımcıların bu dönemde artan endişeleri, doğrudan BİST’i etkileyen gündem olaylarını oluşturmaktadır. Büyük endişelerin, gerginlik, sessizlik ve telaşın hâkim olduğu ortamda ise kimse ticari işlem yapmak istemez. Paralel yapı ve buna bağlı olarak yapılan darbe girişiminin sonuçlarında BİST 100 endeksine yaklaşık yüzde 7 lik bir kayıp şeklinde yansımıştır.

Borsaların olaylardan nasıl etkilendiğini anlayabilmek için, yatırımcıların psikolojisinin ele alınması gerekir. Belirsizlikler, gerginlikler ve anlaşmazlıkların hüküm sürdüğü ortamlar ticaretin yapılabilmesi için uygun olmayan ortamlardır.

Bundan dolayı hiç kimse işlem yapmak istemeyecek ve var olan pozisyonlarının da uygun olması durumunda ivedilikle ayrılmak isteyecektir.

Borsa gündem olaylarından nasıl etkilenir sorusuna en kısa ve kesin cevap olarak, yatırımcıların psikolojisinden direk etkilenir diyebiliriz (<http://borsanasiloyunanir.co/borsa-gundem-olaylarindan-nasil-etkilenir/>).

1.4.2. Borsa Tahmini ile Yapılan Çalışmalar

White (1988: 456), tasarlamış olduğu basit bir YSA ile IBM'nin hisse senetlerinin değerini hesaplamaya çalışmıştır. White geliştirdiği bu sistem ile önceden tahmin edilemeyeceği düşünülen fiyatların aktif pazar hipotezine karşıt tahmin edilebilmesine dair bir kanıt bulamamıştır ama oluşturulacak YSA sisteminde giriş seçeneklerinin artırılması ve veri toplama sürecinin değiştirilmesi ile daha başarılı sonuçlar elde edeceği sonucuna varmıştır.

Grudnitski ve Osburn (1993:639), "Random Walk (RW)" öngörüsünün gelecek fiyatların tahmini için gerçekçi olmadığını, bu öngörünün nonlineer ilişki ve gürültüleri gidermek için bir araç olduğunu kabul etmişlerdir. Yapmış oldukları bu çalışma ile non-lineer olarak tasarlanan YSA'larla S&P 500 endeksi ve altın fiyatlarının tahmin edilebileceğini ispatlamışlardır.

Liu ve diğerleri (2004: 241), borsa trendini tahmin etmek için çok- katmanlı ve ileri- beslemeli bir YSA sistemi kullanmışlardır. Borsa olarak Tokyo Stock Exchange Prices Index (TOPIX) ve girdi olarak da TOPIX de kullanılan teknik analiz parametrelerini seçmişlerdir. Geliştirdikleri sistem günlük olmayıp, günlük olarak sinyal vermek yerine, belirli analizler yaparak kırılma anlarında "AL" ya da "SAT" sinyalleri verir ve sistemin performansı incelendiğinde sistemin çok başarılı olduğu görülür.

Chan ve Teong (1995: 5), yaptıkları çalışmada, teknik analiz tekniklerinin lineer yapısından ve borsanın aktif yapısı ile arasında uyumsuzluk olmasından dolayı teknik analiz yöntemleri ile birlikte kullanılan bir YSA sistemi önermişlerdir. Bu çalışma da hedef, en çok kullanılan analiz teknikleri ile basit bir YSA sistemi oluşturarak, teknik analiz değerlerinin performanslarını iyileştirmek ve kötü bir çalışma yapısı ile zarar eden sistemleri kar eden sistemlere dönüştürmektir. Sistemin veri seti olarak Alman

Markı kullanılıp, oluşan gereksiz karmaşıklıkları önlemek için 1 Ocak 1992'den 30 Mart 1995'e kadar olan süre içindeki yüksek ve alçak değerler ile kapanış değerleri alınarak ileri-beslemeli bir algoritma tercih edilmiştir. Sonuçlar gözlemlendiğinde kar-zarar oranları ve grafikler açık bir şekilde YSA kullanılarak oluşturulan modellerin, YSA kullanılmadan oluşturulan modellere göre daha başarılı olduğu tespit edilmiştir. Kısacası piyasalar bir trende sahip olmadığı zamanlarda teknik analizler kötü sonuçlar verebilir ve bu durum da ancak YSA modellerinin kullanımı ile çözülebilmektedir.

Schierholt ve Dağlı (1996: 72-78, 24-26) çalışmalarını, YSA altyapısını kullanarak hazırlamışlardır. Yaptıkları çalışma da, S&P 500 indeksi irdelenmiş, çok-katmanlı yapı ve istatistikî sinir ağı kullanılarak modeller oluşturmuşlardır. Bu çalışmanın amacı en başarılı tahmin yapmak değil, en iyi kazancı sağlayabilmektir. Her bir kategorinin eğitim örüntüsü ile olasılık yoğunluk fonksiyonu hesaplanmış ve Parza'nın ağırlıklandırma fonksiyonu olan Gauss fonksiyonu kullanılmıştır. Uygulamanın sonuçlarına bakılarak istatistikî sinir ağları modelinin çok-katmanlı modele göre daha iyi ve daha başarılı olduğu sonucunu çıkarmak ve böyle bir genelleme yapmak erken bir kanı olmasına rağmen sinir ağlarında kullanılan diğer modeller göre daha avantajlı olduğu söylenebilir.

McNelis (1996:43), Brezilya borsasını incelediği çalışma da YSA'ların hisse senedi tahminlerinde kullanılabileceği ile ilgili sonuçlar bulmuştur.

Atiya vd. (1997: 2112-2115) geliştirdikleri YSA modelinde, şirketlerin temettü değerlerini, yıllık gelirlerini ve satışlara ait finansal bilgileri yardımcı verilerek olarak kullanılmışlardır. Hisse senetlerini en çok etkileyen etkenlerden biri olan kazançlar ile ilgi olan veriler üzerinde durularak geliştirilen modelin başarılı olduğu bulunmuştur. Model, S&P500 indeksinde bulunan hisse senetleri ve 1993 yılında bu hisse senetlerine ait veriler kullanılmıştır. Uygulanan testlerin sonucunda, YSA modelinin, Al-ve-Tut modelinden daha iyi sonuçlar verdiği tespit edilmiştir.

Wultrich ve diğ. (1998: 368), yapmış oldukları çalışmada 5 büyük borsa endeksine (ABD- Dow Jones, Japonya- Nikkei 225, İngiltere- FTSE 100, Hong Kong- HIS ve Singapur- SSI) ait günlük açılış fiyatlarını tahmin edebilen çevrimiçi bir model geliştirmişlerdir. Tahminde bulunurken Wall Street Journal'da yayınlanan elektronik haberler kullanılmıştır. Modelde borsa endeksi üzerinde etkiye sahip olabileceği düşünülen ve bir uzman tarafından seçilen yaklaşık 400 anahtar kelime esas alınmıştır. Bu anahtar kelime grupları, son 100 işlem gününün kapanış fiyatları ile haber verileri

eđitim verisi olarak belirlenmiřtir. Haber metinleri kullanılarak oluřan anahtar kelimelerin adetlerine bađlı farklı farklı ađırlıklandırma yontemleri uygulanmiř; Kural Tabanlı, En Yakın Komřu ve YSA gibi ođrenme algoritmaları kullanarak sistem eđitilmiřtir. Eđitilen bu sistem gún iindeki haberler ve endekslerin gún sonunda ki kapanıř fiyatları kullanılarak, bir sonra ki gune ait endekslerin fiyatlarının nasıl deđiřeceđi tahmin edilmiřtir. Kural Tabanlı ođrenme algoritmasının diđer algoritmalara góre daha bařarılı olduđu gızlenmiřtir. Eđitilen bu sisteme, gemiř 100 gúnün fiyat verileri eklendiđi zaman sistemin dođruluk oranının yukseldiđi tespit edilmiřtir. Sistemin dezavantajı ise sınıflandırıcı eđitiminin belirli bir grup anahtar kelimelerin kullanılarak yapılmasından dolayı haber metinlerinde yer alan yeni ve endeks üzerinde dođrudan bir etkiye sahip bir kelime oluřtuđunda sistem bu kelimeyi yok sayar.

Yao ve Tan (2000: 291-296), yaptıkları alıřmada, diđer modellerden farklı olarak, YSA kullanan modellerin esas kriteri olan tahmin dođruluđuna ek olarak kârlılık özelliđini de göz önünde bulundurmuřlardır. Finansal zaman serisine dayalı tahmin de bulunurken tüm verileri eřit bir şekilde ađırlıklandıran modeller, tahmini yapılacak olan güne yakın verileri daha fazla ađırlıklandıran modellere góre düşük bir performans göstermiřlerdir. Bundan dolayı bu model, hem kârlılıđın yüksek olması için hem de tahmini yapılacak güne yakın verilerin ađırlıklarını daha yüksek yapabilmek için geri-yayılmalı ađ eđitimine özđu bir kâr esaslı ve ađırlık etkeni kullanılmıřtır. 260'ı eđitim de ve kalan 20'si de test için kullanılan toplam 280 veri ieren bu sistem de bu veriler modelde yer alan 7 ayrı market indisi iinde incelenmiřtir. Bu sistemin performansı incelendiđinde uygulanan bu model ile tahmin oranının bařarisının arttıđı görülmüřtür.

Chan Man-Chung vd. (2000: 61), yapmıř oldukları alıřma da finansal zaman serilerini kullanarak gelecek deđerleri tahmin etmeye alıřmıřlardır bunun için Eřlenik Gradyan Öđrenim (EGÖ) algoritması ile ođul Lineer Regresyon (LR) yontemlerini kullanılmıřlardır. Shanghai borsasını inceleyerek, sistemin giriř deđerleri olarak normalize řekle getirilen SMA, MACD ve RSI gibi 10 tane teknik analiz indeksi eklenmiřtir. Modele en ilk Rasgele İklendirmeli Eřlenik Gradyan (Conjugate Gradient with Random 42 Initialization - CG/RI) metodu uygulanıp daha sonra sıra ile ođul Lineer Regresyon İklendirmeli Eřlenik Gradyan (Conjugate Gradient with Multi Linear Random Initialization - CG/MLRI) ve Rasgele İklendirmeli En Dik Düşüş (Steepest Descent with Random Initialization - SD/RI) metotları model üzerine uygulanmıřtır. Bu yontemlerin performanslarına bakıldıđında en bařarılı yontem yüzde %73.545 dođru

tahmin oranı ile CG/MLRI, daha sonra yüzde %73.055 ile CG/RI, %72.564 ile SD/RI ve en son yüzde %69.303 ile SD/MLRI olarak bulunmuştur.

Gidofalvi (2001: 8), haber metinlerinin kullanılması ile kısa süre içinde hisse senetlerinin hareket yönünü tahmin edebilen bir model geliştirmiştir. 12 tane şirketin hisse senetlerinin gün içinde sahip olduğu fiyat verileri ve finansal haberlerin yayınlanma zamanlarını göz önüne alarak çalışmalar yapmıştır. Bu çalışmada, gün içinde 10 dakikalık zaman dilimlerinde hisse senetleri fiyatlarının değişim değerleri ile NASDAQ endeksinin değişim miktarını kullanarak bir hesaplama yapmıştır. Haberlerin yayınlanma zamanında bulunan skorların kullanılması sonucunda zaman aralıkları ile haberlerin hisse senedi fiyatları arasında anlamlı bir ilişki olduğunu bulmaya çalışmıştır. Bütün haber metinleri kullanılarak aşağı, yukarı ve sabit olmak üzere 3 sınıf etiketi atanmıştır. Bu etiket atama sürecinden sonra Naïve Bayes sınıflandırma algoritması kullanılarak bir takım deneyler yapılmıştır. Deneylerin sonuçlarına göre hisse senedi fiyatı üzerinde anlamlı değişikliklerin haber metinlerinin yayınlanmasından 20 dakika önce veya yayınlandıktan 20 dakika sonrasında olduğu gözlenmiştir. Fakat bu çalışma da haber metinlerine ön işleme yapılmaması ve öznitelik seçmek için bir takım yöntemlerin kullanılmaması bu çalışmanın eksikliğidir.

Pavlidis vd. (2003: 7- 8), yapmış oldukları çalışmada güdümsüz gruplandırma ve evrensel eğitilmiş YSA modelleri kullanılmışlardır. Bir zaman serisi tahmin yöntemi geliştirilmişlerdir. Bu yöntem, kaotik zaman serisi analizi, gruplama, YSA ve evrimsel hesaplama yöntemlerinin kullanılması ile Japon Yen'i ve Amerikan Dolar'ı, Amerikan Dolar'ı ve İngiliz Sterlin'i arasında gelecekte oluşacak kur değişim oranlarını tahmini etmektedir. YSA modelinde ise ileri-beslemeli model seçilmiştir. Sistemde ki veriler 5 evreli bir sistemden geçirilip gerekli işlemlerin uygulanması ile oluşturulmuştur. Sistemin performansına bakıldığında yaklaşık %80lik bir başarının olduğu görülmektedir.

Lam (2004: 368), yapmış olduğu çalışma da geri-yayıma algoritmasına dayalı bir YSA modeli kullanmıştır. Model özellik olarak 16 finansal durum ve 11 makroekonomik parametre kullanılıp, temettü gelirinin verimliliğini tahmin etmeye çalışmıştır. 1985- 1995 arasında ki süre için geçerli olan S&P bünyesinde olan 364 tane şirketin finansal verileri ve makroekonomik parametreleri, ilgili veri tabanlarından meydana gelmiştir. Bu çalışma da veri üzerine oluşan gürültünün zararını giderebilmek için de bir teknik kullanılmıştır. Bu teknik bağlantı ağırlıklarını eğiterek yapay sinir

ağlarını simgesel sınıflandırma kurallarına çevirmiştir. Sonuçları incelendiğinde YSA modeli ile tasarlanan sistem, en düşük değerlendirme oranını yüksek ölçüde ve istikrarlı bir şekilde sağlayıp, en yüksek değerlendirme oranını sağlayamamaktadır.

Mittermayer (2004: 63) NewsCats isminde bir tahmin sistemi tasarlamıştır. Bu sistem, şirketler ile ilgili basın bültenlerinin yayınlanması ile hisse senetleri fiyatlarının nasıl bir değişim göstereceğini tahmin etmek için geliştirmiştir. Sistem 3 ögeden oluşur. İlki, metin işleme yöntemlerinin kullanılması ile basın bültenlerinden gereken bilgiyi çıkarmaktır. İkincisi, işlenenmiş basın bültenlerini daha önceden belirlenen sınıflara ayırmaktır. Sonuncusu da yapılan bu sınıflandırmaya göre hisse senetleri fiyatlarının hareket yönünü bulmaktır. Çalışma da metin işleme yöntemleri olarak, öznitelik seçimi ve doküman gösterimi kullanılmıştır. Birinci bileşende öznitelik seçimi için terim sıklığı, ters doküman sıklığı ve terim sıklığı –ters doküman sıklığı gibi teknikler esas alınmıştır. Doküman gösteriminde de ikili gösterim veya tf , idf ve $tf \times idf$ teknikleri kullanılmıştır. İlk adımda işlenmiş olan basın bültenleri ikinci adımda iyi veya kötü olarak iki sınıfta incelenip, sınıflandırılmış basın bültenlerini kullanıp Destek Vektör Makineleri tekniği ile metin sınıflandırıcıları oluşturmuştur. Yeni gelen basın bülteni için de sistemin ilk adımında yer alan metin işleme yöntemi uygulanmış, ikinci adımda basın bültenin iyi haber mi kötü haber mi olduğu kararı verilmiş, son adımda da bültenin hisse senedi fiyatının hareket yönüne dair bir sinyal oluşturulmuştur. Geliştirilen bu sistem basın bültenlerin yayınlanmasından sonra rastgele bir şekilde alım satım yapan yatırımcı için anlamlı bir başarı sunmuştur.

Zhora (2005: 2549-2554), çalışmasında bir sonra ki günün piyasa değerini tahmin edebilmek için Rasgele Alt Uzay Sınıflandırıcısı (RAS) kullanmıştır. RAS; karmaşık (kompleks), örtüşebilen ve çok boyutlu sınıf dağılımları için çözümler sunabilen, yüksek performanslı sahip bir sinir ağı sınıflandırıcısıdır. Sistemin veri seti IBM'in 1998–2000 yılları arasındaki hisse değerlerini eğitim verisi olarak, 2001 yılındaki hisse değerleri de test verisi olmak üzere deneye tabi tutulmuştur. İlk adımda sistemin girdilerini oluşturan ham fiyat ve hacim değerleri sistemin girişine verilmiştir. İkinci adımda, teknik analiz yöntemlerini kullanarak ham değerler normalize edilmiştir. Üçüncü adımda ise girdi verisine bağlı olarak, bir sınıflandırıcı yöntemi kullanmıştır. Sistemin performansı incelendiğinde yaklaşık %58'lik bir başarı oranı bulunmuştur ve bu oran da genel olarak sistemin başarılı bir şekilde hazırlandığını göstermektedir.

Xiong vd. (2005: 1032), yapmış oldukları çalışmada, Radyan Taban Fonksiyonu (Radial Basis Function – RBF) kullanarak bir YSA modeli tasarlamışlardır. RBF, 3 katmandan oluşan, ileri-beslemeli bir YSA sistemidir. Birinci katman, sinyal kaynak düğümlerinden meydana gelen giriş katmanı, ikinci katman düğüm sayılarının belirli bir amaca göre belirlendiği gizli katman ve son katman da çıkış katmanıdır. Giriş katmanı ile gizli katman arasında oluşan dönüşüm lineer bir yapıda değilken gizli katman ile çıkış katmanı arasında oluşan dönüşüm lineerdir. Geliştirdikleri sisteme girdi olarak bir ülkenin 2002 – 2003 yılları arasındaki vergi gelirlerini vermişlerdir. Sistemin performansına bakıldığı zaman, sistem genel olarak başarılıdır ve ortalama hatanın % 8'den düşük olması nedeni ile de sistemin piyasa tahminlerinde kullanılmak için uygun bir yöntem olduğu görülmektedir.

Li ve Xiong (2005: 3475), geliştirdikleri sistem diğer YSA modelleri ile aynı şekilde çalışır ancak onlardan farkı Bulanık Sinir Ağın'da (BSA) kullanılırlar. YSA modellerinin birçok yararı olsa da niteliksel bilgi karşısında yetersiz olması ve gerçek hayatta ki uygulamalarda problemlere neden olduklarından dolayı ve bu nedenleri aşabilmek için BSA kullanmışlardır. Sistemin veri seti olarak Shanghai borsası seçilmiş ve 5 katmanlı Uyarlamalı Ağ-Temelli Sonuç Çıkarım Sistemi geliştirilen sistem üzerinde denenmiştir. Sistem de ki bağıl hata %1'den az ve sistemin doğru eğilim değeri de %90'dan fazla bulunmuştur. Bu sonuçlara göre model, YSA'nın zaman serisi analizine başarı bir şekilde uygulandığını gösterir.

Majhi vd. (2006: 1-5), yaptığı çalışma da tek katmanlı bir YSA modelini kullanarak döviz kurunu tahmin etmeye çalışmıştır. Tek katmanlı bir model kullanmasının nedeni, tek katmanlı sistemlerin karmaşık olmaması ve çok katmanlı sistemlerle aynı performansa sahip olmasıdır. Sistemde kullanılmakta olan veriler orijinal şekilleri ile kullanılmamıştır. Sistemin daha iyi çalışabilmesi için veriler üzerinden bazı istatistiksel nitelikler seçilip, bunlar normalize edilmesi oluşturulan halleri sisteme girdi olarak verilmiştir. Hindistan Rupisi, İngiliz Sterlin'i ve Japon Yen'i incelenerek, trigonometrik fonksiyonlar kullanılarak bir ay sonraki değerleri tahmin edilmeye çalışılmıştır. Sistemin çalışması incelendiğinde Rupisi %2,3 hata oranı, Sterlin %2.7 hata oranı ve Yen %3,1 hata oranı ile tahmin edilmiştir. Bu sonuçlara bakılarak sistemin başarılı tahminlerde bulunduğu görülmüştür.

Toraman (2008: 32) yaptığı çalışma da demir çelik sektöründe bulunan Erdemir A.Ş ve Kardemir A.Ş hisse senetlerinin 02- 31 Ocak 2008 (22 gün) tarihlerindeki

günlerin kapanış değerlerini kurmuş oldukları YSA modelleriyle tahmin etmeye çalışmışlardır. Modelde kullanılan eğitim ve test verileri için 2003-2007 yılları arasındaki verileri kullanmıştır. Tahmin sonuçlarına göre Erdemir A.Ş hisse senedi fiyatında hata oranı %1,42, ortalama mutlak hata 12 kuruş olarak bulunmuştur. Kardemir A.Ş hisse senedi fiyatında ise hata oranı %1,69, ortalama mutlak hata 2 kuruş olmuştur. Fiyatın yön tahmini Erdemir A.Ş için %95 oranında, Kardemir A.Ş için %90 oranında doğru tahmin edilmiştir.

Bahadır (2008: 76) yapmış olduğu tez çalışmasında uzman sistemler ile YSA modeli kullanılarak, NYSE'de bulunan ETF'lerin gelecek değerlerini tahmin etmeye çalışmıştır. Elde edilen sonuçlara göre iki model de B&H modelinden daha başarılıdır. YSA modeli de uzman sistemlerden daha başarılı sonuçlar vermiştir. 70 ETF yıllık olarak B&H modeli ve YSA modeli ile karşılaştırıldığında iki model arasında büyük farklar olduğu görülmüştür. 70 ETF'den 69'unda YSA modeli, B&H modelinden daha başarılı sadece bir ETF'de B&H modeli %5 daha başarılı bir sonuç vermiştir.

Carvalhall ve Riberio (2008: 92-110), Latin Amerika ülkelerinin hisse senedi endeksleri için 1994-2006 yıllarında ki günlük veriler ile YSA, ARIMA, GARCH ve RW modellerini kullanarak fiyatları tahmin etmeye çalışmışlardır ve YSA modeli ile daha başarılı sonuçlar elde edildiğini bulmuşlardır.

Birgül ve Bertan (2009: 34) BİST100 endeksi tahmini için farklı makro değişkenler kullanarak ileri beslemeli YSA modelleri oluşturmuşlardır. 127 günlük tahmin için endeks yönünü %55,1 oranında bir doğrulukla tahmin etmişlerdir. Buldukları oran, hareketli ortalama tekniği ile bulunan %50,4'lük orandan daha iyi sonuç vermiştir.

Vaisla ve Bhatt (2010: 2108) Hindistan için yapmış oldukları hisse senedi fiyatı tahmini için, mutlak ortalama hata ve hataların karelerinin toplamının ortalaması gibi çeşitli performans kriterlerine göre oluşturulan YSA modelinin regresyon ile oluşturulan modele göre yaklaşık olarak 9 kat daha başarılı olduğu bulunmuştur.

Akcan ve Kartal (2011: 37), BİST'te yer alan ve sigorta sektörünü oluşturan 7 şirketin 15 gün sonra ki hisse senedi fiyatını kurmuş oldukları YSA modeli ile tahmin etmeye çalışmışlardır. 15 gün sonra ki tahmin için hata oranı %0,85 ile %2.36 arasında değişirken, mutlak hata 2 kuruş ile 27 kuruş arasında değişmiştir.

Kara vd. (2011: 5318), kurmuş oldukları YSA modelleri ile BİST100 endeksinin yönünü tahmin etmişlerdir. Çıkış vektöründe 0 azalışı ifade edecek ve 1 de

artışı ifade edecek şekilde belirledikten sonra BİST 100 endeksinin yönünü %75,4 oranında bir doğrulukla tahmin etmişlerdir.

Aghababaeyan (2011: 16), kurduğu YSA modeli ile Tahran borsasında yer alan bir hisse senedinin fiyatını aylık olarak tahmin etmiştir. Bu çalışmada elde edilen tahmin ile gerçek sonuç arasında ki korelasyon 0,97 olarak bulunmuş ve fiyatın yönü de %83 oranında bir doğrulukla tahmin edilmiştir.

Ticknor (2013: 5505), yaptığı çalışma da Bayes eğitim fonksiyonuyla eğittiği YSA modeli ile Microsoft Corporation şirketinin hisse fiyatını yüzde %1,06 hata ile Goldman Sachs şirketinin hisse fiyatını da yüzde %1,33 hata ile günlük olarak tahmin etmiştir.

Gündüz (2013: 49-51) yaptığı çalışma da belirli bir internet sayfasında yayınlanan ekonomi haberlerinin kullanılması ile BIST 100 endeksinin günlük açılış fiyatını tahmin etmeye çalışmıştır. Çalışmada kullanılan ekonomi haberleri Borsa İstanbul'da yer alan şirketlerin resmi bir şekilde bildirimlerinin yayınladığı Kamu Aydınlatma Platformu (KAP) internet sayfasından ve iki finans sayfasından toplanmıştır. Haber metinleri üzerine metin madenciliği algoritmalarının uygulanması ile işlem günlerine ait öznitelik vektörleri belirlenmiştir. Bu öznitelik verileri ile BIST 100 endeksine ait günlük açılış fiyatlarındaki değişimler gözlenmiştir.

Shah vd. (2014: 2114) kurmuş oldukları çeşitli YSA sistemleri ile Bombay Stock Exchange (BSE) endeksini tahmin etmeye çalışmışlar ve bu sistemlerin tahmin sonuçlarını karşılaştırmışlardır. Oluşturulan YSA modellerinden çok katmanlı ileri beslemeli olanlar (ÇKİB) geri beslemeli olanlara göre daha başarılı olmuştur. ÇKİB modelinde geri yayılım algoritması, radyal bazlı algoritmaya göre daha iyi sonuçlar vermiştir. Sonuç olarak en iyi tahminde bulunan YSA ÇKİB modelinden geri yayılım algoritması olmuştur.

Patel ve Yalamelle (2014: 13760) yapmış oldukları çalışma ile Hindistan borsasında LIX15 endeksindeki 15 hisse senedinin fiyat yönünü %51,06 doğruluk ile tahmin etmişlerdir. Bu doğruluk oran rastlantısal bir değer değildir. Bu sonuç oluşturulan YSA'nın yapısını, parametrelerini, veri sayısını ve değişkenlerin seçimi gibi etkenlerin bir sonucu olarak yorumlanmaktadır.

Özer (2015: 63) tarafından yapılan çalışma da seçilen bir kaynaktan ekonomi haberlerinin toplanması, Bagof- Words (ön tanımlı kelimeler) tekniği ile haberleri olumlu veya olumsuz olarak ayrıştırabilmek için bir uygulama geliştirilmiştir.

Ayrıştırma işleminden sonra bu ayrıştırma işleminin başarı yüzdesi ele alınmıştır. BİST hareketlerinin olumlu ve olumsuz haberlerden etkilenmesine dair ve geleceğe dair bir tahmin de bulunup bulunulmayacağını dayalı bir portföy yönetim stratejisi önermiştir.

1.5. Veri Madenciliği

1.5.1. Veri Madenciliği Tanımı

Veri Madenciliği, veri tabanlarında bilgi keşfi (VTBK), veri örüntülerinin bulunması için akıllı yöntemlerin uygulanmasıdır. (Camacho ve Borges 2005). Bu keşif, bilgi teknolojilerinin doğal yapısının bir sonucu olarak görülmektedir (Han ve Kamber 2006).

“Veri Madenciliği (VM) geçerli ve etkin olan ama daha önceleri bilinmeyen, bilgilerin büyük veri tabanlarından elde edilmesi ve bu elde edilen bilgilerin daha sonraki iş kararlarını almada kullanılmasını kapsar (Cabena, 1998).”

“Bilgi keşif sürecinde bir adım olan VM, verilerde ki örüntüleri bulan algoritmaları içerir. Bulunan bilgi sonrası için bir tahmin yapabilmek, sınıflandırma modeli kurmak, birliktelik kurallarını belirlemek, olan bir modeli yenilemek veya model üzerinde yapılmış olan madencilik çalışmasının özetini çıkarmak için kullanılabilir (Fayyad, 1996).”

“Büyük miktarda bulunan veri üzerinde anlamlı örüntü ve kurallar bulmak için çeşitli yöntemler kullanarak veriyi inceleme ve analiz etme işlemidir (Berry ve Linolf, 2000).”

“VM, organizasyonların sahip oldukları veri tabanlarındaki en önemli bilgilere odaklanabilmesine imkan verir. Bununla birlikte yöneticiler gelecekte oluşabilecek eğilim ve davranışları önceden tahmin edebilmek için daha iyi kararlar alabilirler (Chopoorian, 2001).”

“VM, bilinen verilerden üstü kapalı, çok belli olmayan veya daha önceden bilinmeyen fakat belirli bir potansiyele sahip olan kullanışlı bilgilerin çıkarılmasıdır. Kümeleme, veriden özet çıkarma, değişiklikleri belirleme sapmaları tespit etme gibi belirli türdeki teknik çözümleri kapsar (Grossman, 2001).”

“VM, istatistiksel yöntemler olarak kabul edilebilir. Fakat veri madenciliğinin bilinen istatistik yöntemlerden birkaç açıdan farkı vardır. VM de amaç, veriden kolay bir şekilde mantıksal kurallar çıkarmak veya verileri görsel sunumlara çevrilebilecek

nitel modelleri oluşturmaktır (Dönmez, 2008).”

“VM, kullanıcının verilerden açık ve faydalı sonuçlar çıkarabilmesi için, büyük miktardaki verilerden başlangıçta fark edilmeyen, bilinmeyen bazı ilişki ve kuralların keşfedilmesi için, verinin seçilip incelenmesi ve veriden modeller çıkarılmasını kapsayan süreçtir (Giudici,2003).”

“VM, büyük miktarlarda bulunan verilerden gelecek için tahminlerde bulunmaya yardımcı olacak yararlı bağlantı ve kuralların aranması ve analizidir. Bunun yanı sıra veri madenciliği, büyük miktarda ki verilerin içindeki ilişkileri araştırarak bunların aralarındaki bağlantıları bulmada kullanılan bir veri analizi tekniğidir. (Kalikov, 2006)”

VM, geçmiş verileri kullanarak, gelecekte oluşabilecek olayları önceden tahmin edebilmektir. Bundan dolayı veri madenciliği için 3 anahtar teknolojinin beraber kullanılması yapılacak işlemlerin kalitesi açısından önem taşımaktadır. Bu 3 teknoloji şöyledir; (<https://ahmetsedef.wordpress.com/tag/veri-madenciligi-sureci/>)



Kaynak: <https://ahmetsedef.wordpress.com/tag/veri-madenciligi-sureci/>

Şekil 1. Veri Madenciliğinde Kullanılan 3 Anahtar Teknolojisi

1.5.1.1. Bilgi işlem gücü

Verilerin işlendiği sunucularda işlemler CPU tarafından yapılır. Seri işlemleri yapabilmek için tasarlanmış yongalardır. CPU’lar paralel işlemleri seri işlemler kadar iyi yapamazlar. Paralel işlemleri gerçekleştirebilmek için son dönemlerde GPU ya da GPU üzerinde GPGPU programlama teknikleri geliştirilmiştir. Teknolojinin gelişimi ile bileşenlerin (komponent)maliyetleri azaldığı için güçlü bir alt yapı oluşturma maliyeti de azalmaktadır. Geçmişte verileri depolama saklayabilme problemleri yaşanırken günümüzde bu problem çözülmüştür. Öyleki son zamanlarda “Büyük veri (Big data)” kavramı ortaya çıkmış ve bu konu ile ilgili yapılan çalışmalar çok hızlı bir şekilde artmaya devam etmektedir(<https://ahmetsedef.wordpress.com/tag/veri-madenciligi-sureci/>).

1.5.1.2. Data tutma ve yönetme yetkinliği

Teknolojinin gelişmesi ile veri depolama maliyetleri azalmış ve bununla birlikte orta büyüklükteki bütün şirketler verilerini sistemli bir şekilde tutmaya başlamışlardır. Bu şirketler depoladıkları verileri etkin bir şekilde kullanamıyor olsalar bile çok büyük miktarda veri tutmanın daha iyi olduğunu düşünürler. Çünkü ihtiyaç duydukları verinin ellerinde bulunmasını isterler. Büyük ölçekteki şirketler ise sadece kullanacakları veriyi veya ürettikleri veriyi veri depolama sistemlerinde tutarlar. Sürekli bir gelişim içerisinde olan veri depolama sistemleri, verileri daha nicelikli halde tutarak, kullanıcı için verilerin yönetimini kolaylaştırır (<https://ahmetsedef.wordpress.com/tag/veri-madenciligi-sureci/>).

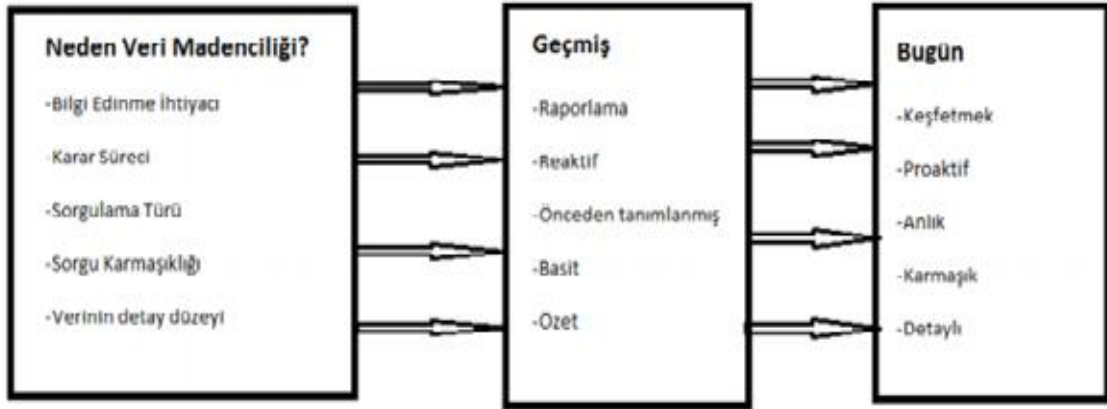
1.5.1.3. Analitik araçların gelişmesi

Teknolojinin gelişimi istatistiksel ve analitik yöntemlerin büyük veriler üzerinde uygulanmasına imkân sağlamıştır. Veri madenciliği sürecinde yapılan veri madenciliği işlemleri manuel olarak yapılır. Makine öğrenmesi yeni bir kavram olmamasına rağmen, istatistiksel yöntemlerin bir parçası olduğu çoğu zaman unutulmaktadır. VM'nin amacı verilerden faydalı sonuçlar çıkarmak ve sorunları çözmek olduğu için bilinen yöntem ve teknikleri geliştirmek amacıyla yeni veri madenciliği teknikleri incelenmekte ve geliştirilmektedir. Yeni yöntem ve tekniklerin geliştirmesi hem özel sektörlerin öncülüğünde hemde ortak bir geliştirme sisteminde açık kaynak kodlu yazılımların geliştirilmesi ile yapılmaktadır (<https://ahmetsedef.wordpress.com/tag/veri-madenciligi-sureci/>).

1.5.2. Veri Madenciliğinin Önemi

Teknolojinin gelişmesi ile daha ucuz bir şekilde üretilen veri ambarlarının hacimleri artmış ve kullanımları da yaygınlaşmıştır. Eski tip veri tabanlarının hacimleri küçük olduğu için veriler özet şeklinde saklanırken günümüzde kullanılan veri tabanlarının hacimlerinin büyümesi ile veriler daha detaylı olarak saklanmaktadır. Saklanmış olan bu veriler geçmiş ile ilgili rapor çıkarma da kullanılır. Özet verilerin boyutları küçük olduğun için herhangi bir çaba gerektirmeden bu veriler üzerinden bağlantılar kolaylıkla bulunur. Ancak günümüzde daha detaylı bir şekilde saklanabilen verilerden anlamlı bilgi, bağlantı ve ilişkiler bilgisayar algoritmaları kullanılarak tespit edilir. VM algoritmalar yardımı ile anlamlı bilgiler keşfetme sürecidir. Verilerde

detayların olması birden fazla değişken ve duruma neden olmasından dolayı basit yapısını kaybetmiş, daha karmaşık bir hale gelmiştir. Özet şeklinde tutulan veriler üzerinde sorgulama yapmak değişkenlerin sayısının az olmasından dolayı daha kolay olurken, değişken sayısının fazla olduğu durumlarda yani verilerin detaylarının fazla olmasından dolayı sorgulama yapmak daha karmaşık olur. Geçmişten günümüze bilginin önemindeki değişiklik ve bu değişikliklerin veri madenciliğine etkileri şekil 2.2’ de gösterilmektedir.



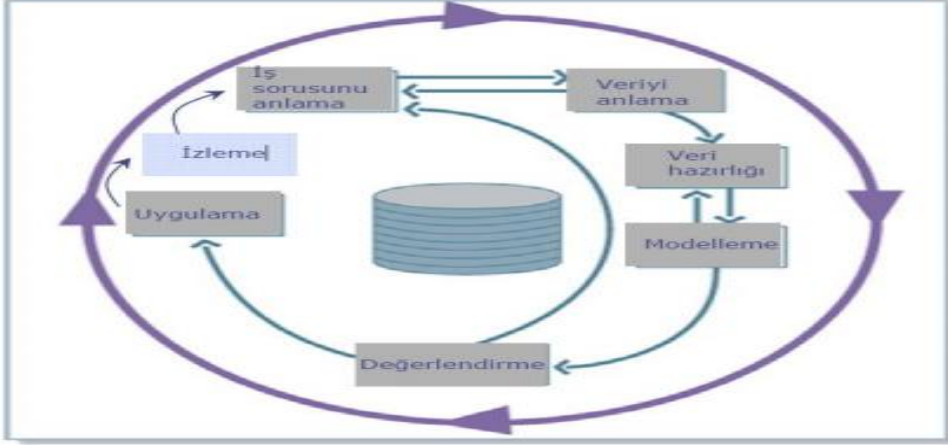
Kaynak: Özmen,2015: 7.

Şekil 2. Veri Madenciliğinin Etkileri

1.5.3. Veri Madenciliği Süreci

VM süreci, projenin yaşam sürecine genel bir bakış getirir. Proje de birbirini takip eden aşamalarını, her aşamanın kendine ait görevlerini ve bu görevler arasındaki ilişkileri kapsar. Bu tanımlama aşamasında bütün ilişkileri belirlemek mümkün değildir (Ahi, 2015: 24).

VM proje süreci belirli aşamalardan oluşur. Aşamalar arasında katı bir sıra vardır. Aşamalar arasında sürekli geriye ve ileriye doğru bir hareketlilik söz konusudur. Bu hareketlilikler her aşamanın sonunda bir sonraki aşama da hangi işlerin yapılacağına bağlıdır. Oklar, aşamalar arasında sıklıkla oluşan bağımlılıkları gösterir.



Kaynak: <http://mustafaakca.com/crisp-nedir/>.

Şekil 3. CRISP- DM veri madenciliği süreci

1.5.4. Veri Madenciliği Kullanılan Yöntemler

1.5.4.1. Sınıflandırma

Veri madenciliğin de en çok kullanılan alan sınıflandırmadır. Daha önceden oluşturulan veri tabanının bir bölümü eğitim amaçlı kullanılması ile sınıflandırma kuralları oluşturulur. Bu oluşturulan kurallar ile yeni bir durum olduğunda nasıl bir karar verileceği belirlenir.

Sınıflandırma da temel olarak iki adım bulunur. Bunlardan birincisi gözlenmiş verilerdir ve bunlar üzerinde çeşitli sınıflandırma algoritmalarının uygulanması ile sınıflandırma kuralları oluşturulur. İkinci adım ise oluşturulmuş sınıflandırma kuralları yeni veriler üzerinde uygulanarak sınıflandırma yapılır. Tahminleyici modeller de sürekli verilerin alınması ile oluşturulan kurallara göre sonuçlandırılır. Sınıflandırma da kullanılan algoritmalarından bazıları Karar Ağaçları (Decision Tree), Yapay Sinir Ağları (Neural Networks), K-en Yakın Komşu (K- Nearest Neighbour), Genetik Algoritmalar, Naive Bayesian Sınıflandırma, Bellek Tabanlı Nedenleme (Memory Based Reasoning) ve Hatayı Geri Yayma'dır (Backpropagation). Ayrık nitelik değerlerini tahmin etme de sınıflandırma ve sürekli nitelik değerlerini tahmin etme de ise öngörü yapılır. (Kumdereli, 2012: 70)

Eğitim Verisi

Müşteri	Borç	Gelir	Risk
Ali	Yüksek	Yüksek	Kötü
Ayşe	Yüksek	Yüksek	Kötü
Fatma	Yüksek	Düşük	Kötü
Fuat	Düşük	Yüksek	İyi
Ece	Düşük	Düşük	Kötü
Ayla	Düşük	Yüksek	İyi



Sınıflandırma algoritması



Sınıflayıcı Model
EĞER Borç=YÜKSEK ise Risk=Kötü;
EĞER Borç=DÜŞÜK Ve Gelir=DÜŞÜK ise RİSK=KÖTÜ;
EĞER Borç=DÜŞÜK Ve Gelir=Yüksek ise RİSK=İYİ;

Kaynak: <http://ab.org.tr/ab13/bildiri/175.pdf>

Şekil 4. Sınıflandırma model kurma süreci

Sınıflandırma kuralları test verileri kullanılarak belirlenir. Bu kurallar daha sonra test verileri kullanılarak sınanır. Örneğin Ali adında yeni bir banka müşterisinin kredi isteminde olduğunu kabul edelim. Bu yeni gelen müşterinin risk durumunu bulabilmek için önceden bulunan verilerden oluşturulan karar kuralları doğrudan uygulanır. Bu müşteri için Borç=Düşük, Gelir=Yüksek olduğu biliniyorsa risk durumunun Risk=İYİ olduğu sonucuna ulaşılır.

Test Verisi

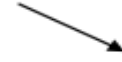
Müşteri	Borç	Gelir	Risk
Cüneyt	Yüksek	Düşük	Kötü
Fatih	Düşük	Yüksek	İyi
Gökhan	Düşük	Düşük	Kötü
Tarık	Yüksek	Yüksek	Kötü



Sınıflayıcı Model
EĞER Borç=YÜKSEK ise Risk=Kötü;
EĞER Borç=DÜŞÜK Ve Gelir=DÜŞÜK ise RİSK=KÖTÜ;
EĞER Borç=DÜŞÜK Ve Gelir=Yüksek ise RİSK=İYİ;



Müşteri	Borç	Gelir	Risk
Ali	Düşük	Yüksek	?



Risk=İYİ

Kaynak: <http://ab.org.tr/ab13/bildiri/175.pdf>

Şekil 5. Test Verisi Üzerinde Sınıflandırma Kurallarının Uygulanması

Yukarıda uygulanan testlerin sonucunda ulaşılan modelin doğru olduğu kabul edilirse, bu model gelecek veriler üzerinde de uygulanır. Sonuçta ulaşılan model mevcut olan ya da olası müşterilerin gelecekte oluşabilecek kredi talep risklerini belirlemek için kullanılır.

Sınıflandırma Yöntemlerini aşağıdaki gibi listelemek mümkündür;

1. İlişki tabanlı sınıflandırıcılar (association-based classifier)
2. Karar ağaçları (Decision tree)
3. K- en yakın komşu yöntemi (K- nearest neighbor method)
4. Yapay sinir ağları (Artificial neural networks)
5. Genetik algoritmalar (Genetic algorithms)
6. Navies Bayes (Bayes classifier)

1.5.4.1.1. İlişki Tabanlı Sınıflandırıcılar

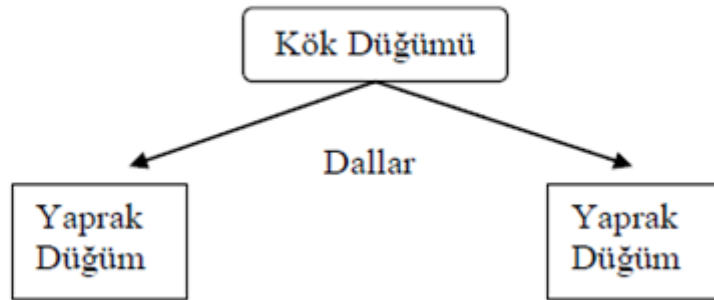
Bu yöntemde temel amaç, oluşan veri setinde ilişkilerin ortaya çıkarılması ve bu ilişkileri kullanarak verilerin sınıflandırılmasıdır. Bu yöntem iki fazdan oluşmaktadır. Öncelikle büyük veri setindeki tekrarlayan verileri bularak ortaya çıkarmak, ikincisi ise bu verileri kullanarak bir model oluşturmaktır. Daha sonra ise bu modeli kullanarak verileri uygun sınıflara ayırmaktır.

1.5.4.1.2. Karar Ağaçları

Karar ağaçları, bir problemin sınıflandırma modelini oluşturmak için uygulanan veri madenciliği yöntemidir.

Karar ağacı yöntemini uygulayabilmek için veri setindeki değişkenler, hedef değişken ve kestirici değişken olmak üzere iki sınıfa ayrılır. Hedef değişken sınıf etiketlerinin yer aldığı değişkenlerdir. Kestirici değişken ise veri setindeki hedef değişkenlerin değerini tahmin etmek için kullanılan değişkenlerdir.

Bir karar ağacının genel olarak düğümlerden (node) ve bu düğümleri birbirlerine bağlayan dallardan meydana gelir. Şekil 6'da olduğu gibi bir karar ağacının kök düğümü başlangıç noktası olarak kabul edilir. Kök düğümün dallanması ile oluşan düğümlerde dallardır. Bir karar ağacının bitiş noktası da yaprak adı verilen yapılardır.



Kaynak: Atasoy, 2015: 17

Şekil 6. Bir Karar Ağacının Temel Yapısı

Bir karar ağacında kategorik değişkenler ile sürekli değişkenler de bulunabilir. Hedef değişken sürekli ise bu karar ağacı regresyon ağacı, hedef değişken kategorik ise sınıflandırma ağacı olarak kabul edilir. Karar ağaçları genellikle bilişim de veri yapıları alanında, medikal için tanı – teşhis yöntemlerinde, botanik için sınıflandırmada,

psikoloji için karar teorisinde ve ekonomi için de yatırım değerlendirme alternatiflerinde tercih edilir (Atasoy, 2015:20).

Tablo 1. Bazı Karar Ağacı Algoritmaları ve Özellikleri

KARAR AĞACI ALGORİTMASI	ÖZELLİKLER
C&RT	Gini'ye dayalı ikili bölme işlemi mevcuttur. Son veya uç olmayan her bir düğümde iki adet dal bulunmaktadır. Budama işlemi ağacın karmaşıklık ölçüsüne dayanır. Sınıflandırma ve regresyonu destekleyici bir yapıdadır. Sürekli hedef değişkenleri ile çalışır. Verinin hazırlanmasına gereksinim duyar.
C4.5 ve C5.0 (ID3 karar ağacı algoritmasının ileri versiyonları)	Her düğümünden çıkan çoklu dallar ile ağaç oluşturur. Dalların sayısı tahmin edicinin kategori sayısına eşittir. Tek bir sınıflayıcı da birden çok karar ağacını birleştirir. Ayırma işlemi için bilgi kazancı kullanır. Budama işlemi her yapraktaki hata oranına dayanır.
CHAID (Chi-Squared Automatic Interaction Detector)	Ki-kare testleri kullanarak bölme işlemi gerçekleştirir. Dalların sayısı iki ile tahmin edicinin kategori sayısı arasında değişir.
SLIQ (Supervised Learning in Quest)	Hızlı ölçeklenebilir bir sınıflayıcıdır. Hızlı ağaç budama algoritması mevcuttur.
SPRINT (Scalable Parallelizable Induction of Decision Tree)	Büyük veri kümeleri için idealdir. Bölme işlemi tek bir niteliğin değerine dayanır. Tüm bellek sınırlamaları üzerinde nitelik listesi veri yapısı kullanılarak işlem yapar.

Kaynak: Çalış, A., Kayapınar, S. ve Çetinyokuş T., 2014: 7

1.5.4.1.3. K-En Yakın Komşu

Denetimli öğrenme yöntemlerinden biri olan K- En Yakın Komşu tekniği ile sınıflandırma problemleri çözülür. Bu yöntem sınıflandırılmak istenen verilerin öğrenme kümesinde bulunan verilere göre benzerlikleri hesaplanır ve en yakın olduğu düşünülen k verinin ortalamasıyla, belirlenen eşik değere göre sınıflara atamaları yapılır. Önemli olan, her bir sınıfın özelliklerinin önceden net bir şekilde belirlenmiş olmasıdır. Yöntemin performansını k en yakın komşu sayısı, eşik değeri, benzerlik ölçümü ve öğrenme kümesindeki normal davranışların yeterli sayıda olması kriterleri etkilemektedir (Çalışkan ve Soğukpınar, 2008: 122).

1.5.4.1.4. Yapay Sinir Ağları

Yapay Sinir Ağları (YSA) kümeleme veya tahminleme, sınıflandırma için kolay bir şekilde kullanabilen güçlü yöntemlerdir.

YSA, sonuçları bilinen belli bir veri seti üzerine öğrenme algoritmaları uygulanarak eğitilir. Bu eğitim ile yapay sinir ağı içinde bazı ağırlıklar oluşur. Bu

ağırlıklandırmalar kullanılarak yeni veriler işlenir ve bir sonuç üretilir. YSA'ların en olumsuz yönlerinden biri de bu ağırlık değerlerinin nasıl oluştuğunun bilinmemesidir. (Ceran, 2006: 33).

1.5.4.1.5. Genetik Algoritmalar

Genetik algoritmalar (GA), biyoloji bilimi kullanılarak geliştirilen makine öğrenme algoritmalarının en önemlilerinden biridir. Çok boyutlu ve karmaşık arama uzayında bulunan öğelerden en iyisinin hayatta kalması için en iyi olan çözümü arar. GA problemlerinde sonuç tek bir çözümden oluşmaz, sonuç için çeşitli çözümlerden oluşan bir çözüm kümesi oluşturulur. Arama uzayında birden fazla nokta değerlendirildiği için bütünsel bir çözüm elde edilme olasılığı artmaktadır. Oluşturulan çözüm kümesindeki çözümler birbirinden bağımsızdır. Her bir çözüm karmaşık ve çok boyutlu olan arama uzayında bir vektörü temsil eder.

GA problemlerin çözüm süreci bilgisayar ortamına ki çözümleri benzer. Diğer iyileştirme yöntemlerinde olduğu gibi tek bir çözüm kullanmak yerine, bu tür çözüm yöntemlerinden meydana gelen bir çözüm kümesi oluştururlar. Problemin çözümü için oluşturulan çözüm kümesinin genetik algoritma terminolojisindeki adı popülasyondur. Vektör ya da birey olarak adlandırılan sayı dizilerin nüfusları oluşturur. Bireyi oluşturan elemanlara gen denir. Nüfusu oluşturan bireyleri genetik algoritma işlemcileri belirler. Genlerin oluşturduğu yapılar olan kromozomlar üzerinden de hesaplamalar gerçekleştirilir.

GA değişik yapılardaki verileri işleyebildikleri için optimizasyon yöntemlerinde de kullanılırlar. GA, YSA ile birlikte çalışarak başarılı sonuçlar üretirler. GA, yapay sinir ağlarının eğitilmesinde, hafıza tabanlı tekniklerde, birleşim fonksiyonlarını oluşturma gibi çeşitli işlemlerde kullanılırlar.

GA olumlu özelliklerine sahip olmasına rağmen kullanımlarında bazen sorunlar yaşanabilmektedir. Karmaşık sorunları genetik olarak kodlayabilmek bu sorunlardan en bilinenidir. GA kullanımı ile bulunan sonucun en iyi olduğuna ilişkin bir garanti de söz konusu değildir (Gülçe, 2010: 22).

1.5.4.1.6. Navies Bayes

Navies-Bayes (NB), Bayes teorisi temeline dayanarak elde bulunan verileri istatistiksel yöntemleri kullanarak sınıflandırır. NB algoritmaları uygulanabilir olması ve

hızlı hesaplama yapabilmesinden dolayı arařtırmacılar tarafından çok kullanılan bir algoritmadır. Sınıflandırılacak veriler birbirinden bağımsız olarak kabul edilir.

NB algoritmasının uygulanışında bazı kabuller yapılmaktadır. Bu kabullerin en önemlilerinden biri niteliklerin birbirinden bağımsız olmasıdır. Niteliklerin birbirlerini etkilediđi durumlarda, olasılıkları hesaplamak zorlaşacağı için niteliklerin aynı seviyede önemli olduđu varsayılır.

NB algoritması sürekli veriler ile çalışmaz. Sürekli veriler bağımlı ya da bağımsız şekilde kategorik bir forma getirilmelidir. NB algoritmasında, modelin öğrenilmesi aşamasında, her çıktının öğrenme kümesinde kaç kere meydana geldiđini hesaplamaktadır. Bulunan bu deđer, öncelikli olasılık olarak adlandırılır. NB aynı zamanda her bağımsız deđişken / bağımlı deđişken kombinasyonunun meydana gelme sıklığını bulmaktadır. Bu sıklıklar öncelikli olasılıklarla birleştirilmek suretiyle tahminde kullanılır (Ahi, 2015: 44).

1.5.4.2. Kümeleme

Eđiticesiz öğrenme ile gerçekleştirilen kümeleme algoritmasında birbirlerine benzeyen nesnelere gruplandırılır (Akkuş, 2015: 8).

Deđişkenli bir analiz tekniđi olan kümelemenin ilk amacı nesnelere özelliklerine göre gruplamaktır. Nesnelere gruplandırma önceden belirlenen ilkeler göre yapılır. Kümeleme algoritmalarının uygulaması sonucu elde edilen sonuçlar, kümeler içinde homojen, kümeler arasında da heterojen olmalıdır. Rastlantı deđişkeni kümeleme içinde önemli bir yere sahiptir. Bu rastlantı deđişkeninin kümeleme işleminde yeri diđer deđişkenli yöntemlerden oldukça farklıdır. Nesnelere karşılaştırmasında kullanılan nitelikleri ifade eden deđişkenlere rastlantı deđişkeni denir.

Rastlantı deđişkenleri sadece nesnelere karşılaştırma da kullanılan deđişkenleri içerdiđinden dolayı nesnelere sahip olduđu özellikleri belirler. Rastlantı deđişkenleri arařtırmacıların belirlediđi çok deđişkenli bir analiz yöntemidir. Kümeleme, rastlantı deđişkenlerini deney sonuçları ile tahmin edemez. Bundan dolayı kümelemenin amacı rastlantı deđişkenlerini tahmin etmek deđildir. Kümelemede rastlantı deđişkenlerine bađlı olarak nesnelere karşılaştırılması yapılır (Taşkın ve Emel, 2010: 397).

1.5.4.3. Tahminleyici Model

Sonuçları daha önceden bilinen verilerden yola çıkılarak yeni bir model geliştirilmesi ve geliştirilen bu model kullanılarak sonuçları bilinmeyen veri setlerinin

sonuç kümesini tahmin edebilmeyi amaçlamaktadır. Bu modeller sınıflama ve regresyon teknikleridir. Tahmini yapılacak durum sayısal (sürekli) bir değişken ise bu bir regresyon problemidir. Tahmini yapılacak durum kategorik bir değişken ise bu bir sınıflama problemidir (<https://emraharslanbm.wordpress.com/tag/tahminleyici-modelleme/>).

1.5.4.4. Birliktelik Kuralları

Birliktelik kuralları, kümelemenin haricinde ki başka bir diğer anahtar danışmansız veri madenciliği yöntemidir; veri parçalarının büyük veri setleri içerisindeki ilgi çekici birliktelikleri (ilişkiler, bağımlılıklar) bulur. Veri parçaları, bir dış süreç tarafından oluşturulur veya ilişkisel veri tabanları yada veri ambarlarından çekilecek işlemler şeklinde depolanır. Birliktelik kuralları algoritmalarının iyi ölçeklenebilir özelliklerinin olmasından ve biriktirilen verinin sürekli büyümesinden dolayı, birliktelik kuralları veriden bilgiyi çıkarmak için gerekli temel bir veri madenciliği aracıdır. İlgi çekici ilişkilerin keşfedilmesi, karar vermek için iş birimleri tarafından çoğunlukla kullanılan bir bilgi kaynağını sağlamaktadır. İş dünyasında en yaygın olan, benzerlik analizi veya pazar sepet analizi olarak bilinen, birliktelik analizi iki veya daha fazla özellik arasındaki ilişkiyi ölçmek için kuralları açığa çıkarmaya çalışır (Karakuş, 2009: 37).

Pazar sepet analizi yöntemiyle bir müşteri herhangi bir ürün satın aldığı anda sepetinde bulunan diğer ürünler de bir olasılığa göre tahmin edilir. Pazar sepet analizleri ile beraber alınan ürünler belirlenip, mağazalarda bulunan raflar da ona göre yerleştirilirse müşteriler bu ürünlere daha kolay bir şekilde erişebilir. Bu modelin en bilinen temsilcisi “Apriori” algoritmasıdır (Gülçe, 2010: 26).

1.5.4.5. Ardışıklık Örüntüsü

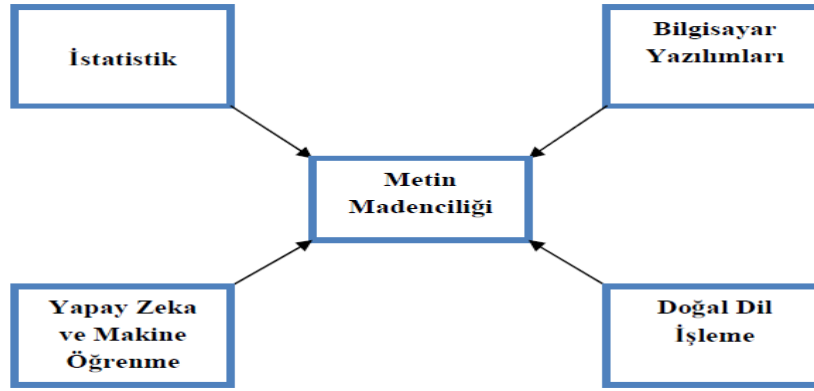
“Ardışık örüntüsünde amaç belirli bir zaman sürecinde sıklıkla gerçekleşen olayların kümelerini bulmaktır (Agrawal ve Srikant, 1995). “

Ardışıklık örüntüsüne örnek olarak; Bir sene içerisinde Orhan Pamuk’un “Benim Adım Kırmızı” romanı satın alan insanların yaklaşık %70’i Buket Uzuner’in “Güneş Yiyen Çingene” kitabını da satın almıştır. Bu tipteki örüntüler perakende satış, telekomünikasyon ve tıp alanlarında da kullanılır (Akbulut, 2006: 25).

İKİNCİ BÖLÜM

2. METİN MADENCİLİĞİ

Metin madenciliği yapısal halde bulunmayan metinlerden ilgi çekici bilgi ve anlamların çıkarıldığı süreç olarak tanımlanır. Metin halindeki bir veriden bilgi elde edebilmek için ise bazı işlemler gerçekleştirilir. Bu işlemler sonucunda yapısal halde bulunmayan metinsel veriler yapısal bir forma dönüşür. Yani metinsel veriler artık veri madenciliği yöntemlerinin uygulanabileceği formata dönüşmüş olur (Hotho, Nurnberger ve Paaß, 2005:19). Şekil 7’de metin madenciliği ile diğer disiplinler arasındaki ilişki gösterilmiştir.



Kaynak: Ergün,2012: 2

Şekil 7. Metin madenciliği ile diğer disiplinler arasındaki ilişki

İstatistik; verilerin toplanması, toplanan verilerin sınıflandırılması, özetlenmesi, analizinin yapılması ve analiz sonuçlarının grafik ve tablolarla sunulması, verilerden anlamlı bilgilerin elde edilmesi ve yorumlanmasıdır. Veri ve metin madenciliğinde ki amaç ile istatistik biliminin amaçları birbirine benzemektedir. Metin ve veri madenciliğinde kullanılan en temel yöntemlerin istatistiksel yöntemler olduğu birçok uygulama ve çalışmada belirtilir (Özmen, 2001: 2).

Metin madenciliği, istatistiksel modeller aracılığıyla metin içinden yararlı ve anlamlı bilgiler ortaya çıkarır.

Yapay zekâ ve makine öğrenmesi de metin madenciliği ve istatistikle ilişkili bir çalışma alanıdır. Metin madenciliği gibi makine öğrenmesi de, analizi yapılacak olan

yapısal veya yapısal olmayan verilerin özelliklerini tanımlayan öznelik değerlerini sisteme gönderir ve daha sonra sisteme gönderilen bu bilgilerden programın bu verilerin örüntüsünü çıkarması veya yapısını öğrenmesi istenir. Makine öğrenmesi tekniği, büyük veri setlerinde bilgiyi daha kolay işleyebilen yöntemlerdir.

Doğal dil işleme (DDİ), yapay zekâ ve makine öğrenmesi alanları metin madenciliği fonksiyonlarının önemli metin madenciliği yöntemleridir. DDİ, doğal insan dilini anlayabilme gibi problemler üzerinde çalışmaktadır. DDİ, metinler gibi doğal dil kullanılarak hazırlanmış dökümlerde bilgisayar programlarının kolaylıkla işlemler yapabileceği sayısal ve simgesel veri formlar gibi daha resmi gösterimlere dönüştürülmesiyle ilgilenir. DDİ'nin amacı, metinlerin sayısal formlara aktarılmasının yanı sıra dilbilimsel, semantik ve içerik kısıtlarını da göz önüne alan doğal dilin anlaşılması ve işlenmesini sağlamaktır. Her ne kadar bu disiplin ile ilgili önemli bir yol kat edilmiş olsa da, dil yapısının karmaşıklığı dikkate alındığında alınan mesafenin çok ileri düzeyde olduğu söylenemez.

Bilgisayar yazılımları ile metin madenciliği arasında özellikle iş dünyası açısından oldukça önemli bir ilişki vardır. Özellik çıkarımı teknikleri ve ağırlıklandırma modellerinin yardımıyla metinlerin yapısal olmayan formdan sayısallaştırılarak yapısal forma dönüştürülmesi, dokümanların sınıflandırılması, kümelenmesi gibi daha birçok alanda bilgisayar yazılımları kullanılmaktadır. Dokümanların içindeki gizli kalmış bilgilerin ortaya çıkarılarak yararlı ve anlamlı modellerin oluşturulmasında, yani metin madenciliğinin tüm süreçlerinde bilgisayar yazılım araçları kullanılmaktadır.

2.1. Metin Madenciliği Çalışma Alanları

Metin madenciliği sırasında genelde aşağıdaki problemlerle ilgilenilir.

Enformasyon Getirimi: İlgilenilen konu hakkında ön araştırmanın yapılp bilginin toplandığı aşamadır. Örnek olarak metin madenciliği web kaynaklı bir veri üzerinden yapılacaksa web sayfasının ismi, adresi veya dosya sisteminde ise o dosyanın isimleri, dizin bilgileri, tarihleri ve kullanıcı bilgileri gibi bilgilerin toplanma aşamasıdır.

DDİ aşaması: Metin madenciliğinin tüm aşamalarında kullanılmasa da genellikle özellik çıkarımı ve metinlerdeki anlamsal bilgilerin çıkarılmasında başvurulur. DDİ, konuşma parçalarının etiketlenmesi ve cümlesel parçalama gibi dilbilimsel işlemler sırasında kullanılır.

Adlandırılmış varlık tanıma: Metin işleme sürecindeki istatistiksel özelliklerin çıkarılmasında kullanılır. Metin içindeki kişi, yer isimleri, sembol ve kısaltmalar bu yöntemle bulunur. Metin madenciliğinin yapılması için metinler her zaman temizlenmiş olmalıdır. Sosyal medyalarından gönderilen mesajlar, telefonlar üzerinden gönderilen SMS mesajları gibi mesajlar genellikle yazım hataları ve kısaltmalar içerir. Metin madenciliği bu ihtimalleri göz önüne alarak işlemlerini gerçekleştirir. Örneğin ‘‘alibey’’ kelimesi, Antalya da bir otel ismi olabileceği gibi bir kişi ismi de olabilir.

Örüntüsü tanımlı varlıkların bulunması: Metin içerisindeki bazı özel bilgiler bazı durumlarda metin madenciliğinin konusunu oluşturabilir. E posta hesapları, telefon numaraları, adresler, kişilik bilgileri veya tarihler gibi bilgiler bazı durumlarda özel olarak alınmak istenebilir. Bu tür durumlarda düzenli ifadeler veya içerikten bağımsız gramerler tanımlanması ile bunlar metin üzerinde çalıştırılır.

Eş Atıf: Bir varlığı işaret eden isim kelime grupları ile diğer terimsel ifadelerin bulunması veya ayrılmasını hedefler.

İlişki, kural, olay çıkarımları: Çeşitli nedenlerden dolayı metin içerisinde bulunan bazı verilerin çıkarılması istenebilir.

Duygu analizi: Metin içerisinde yer alan bazı duygusal ifadelerin çıkarılmasını amaçlar. En çok kullanılanı duygusal kutupsallıktır. Kutupsallık ile bir konu hakkında ki mesaj ve yazıların pozitif veya negatif olacak şekilde sınıflandırılması amaçlanır. Fakat duygu analizi olumlu ve olumsuz olarak ayrıştırma dışında metin içindeki ruh hali, düşünce, davranış gibi daha karmaşık duyguların üzerinde de çalışmaktadır (Şeker,2015:32).

2.2. Metin Madenciliği Kullanım Alanları

Günümüzde metin madenciliğinin çok yaygın bir kullanım alanı bulmaktadır. Bunlara örnek olarak pazarlama, telekomünikasyon, bankacılık, sigortacılık, tıbbi alanlar ve elektronik ticaret ile ilgili alanlar söylenebilir. Bunlar kullanım yerlerine göre aşağıdaki gibi sınıflandırılmıştır. (<http://visualdatamining.blogspot.com.tr/2009/06/veri-madenciliginin-yararlar-ve.html>)

Tablo 2. Metin Madenciliği Kullanım Alanları (1)

Pazarlama	Bankacılık	Sigortacılık
Müşterilerinin satın alma alışkanlıklarının belirlenmesi	Farklı finansal göstergeler arasındaki gizli korelasyonların tespiti	Yeni poliçe talep edecek müşterilerin tahmin edilmesi
Müşterilerin demografik özellikleri arasındaki bağlantıların ortaya konması	Farklı finansal göstergeler arasındaki gizli korelasyonların tespiti	Sigorta dolandırıcılıklarının tespit edilmesi
Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması	Kredi kartı harcamalarına göre müşteri gruplarının belirlenmesi	Riskli müşteri gruplarının belirlenmesi
Pazar sepeti analizi	Kredi taleplerinin değerlendirilmesi	
Müşteri ilişkileri yönetimi		
Müşteri değerlendirme		
Satış tahmini		

Kaynak: <http://visualdatamining.blogspot.com.tr/2009/06/veri-madenciliginin-yararlar-ve.html>

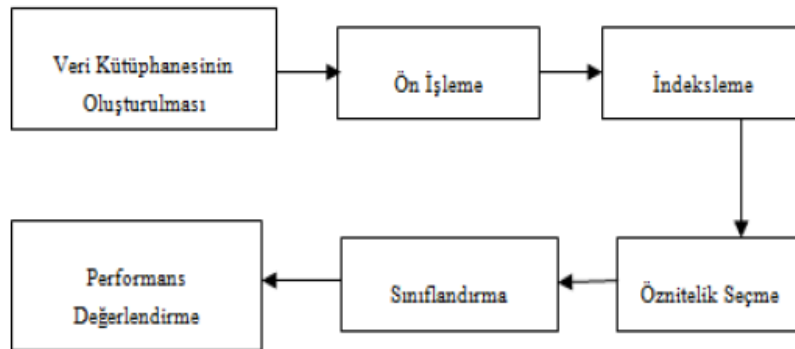
Tablo 3. Metin Madenciliği Kullanım Alanları (2)

Elektronik Ticaret	Telekomünikasyon	Tıbbi Araştırma	Diğer
Saldırıların çözümlenmesi	İletişim ağlarında sorunlu bölgelerin tespiti	DNA içersindeki genlerin sıralarının belirlenmesi	Taşımacılık ve ulaşım
e-CRM uygulamalarının yönetimi	Kaçak hat kullanımlarının belirlenmesi	Protein analizlerinin yapılması	Turizm ve otelcilik
WEB sayfalarına yapılan ziyaretlerinin çözümlenmesi	Kullanıcı davranışlarının belirlenmesi	Hastalık haritalarının hazırlanması	Belediyeler
Kullanıcı davranışlarına göre web sitesinin yenilenmesi	Müşteri davranışlarına göre yeni hizmetlerin sunulması	Hastalık tanıları	Eğitim
		Sağlık politikalarına yön verilmesi	Bilim ve Mühendislik

Kaynak:<http://visualdatamining.blogspot.com.tr/2009/06/veri-madenciliginin-yararlar-ve.html>

2.3. Metin Madenciliği Süreci

Metin işleme süreci genel olarak Şekil 8 'deki gibi çalışmaktadır.



Şekil 8. Metin İşleme Süreci

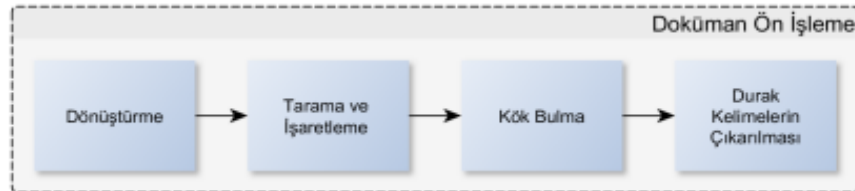
2.3.1. Verilerin Toplanması

Metin işleme sürecinin ilk adımı olan verilerin toplanması aşamasında internet üzerinden veya diğer yöntemlerle ulaşılan veriler toplanarak bir kütüphane veya veri tabanı oluşturulup daha sonra bu veriler üzerinde gerekli sınıflandırma işlemleri yapılmaya başlanır. Tez de veritabanı aşağıdaki işlemlerin uygulanması sonucu oluşturulmuştur.

RSS (Really Simple Syndication): İnternet sitelerinin yayınladığı haber vb. içeriklerin tek bir ortam üzerinden ulaşıp izlenebilmesine imkân tanıyan yeni bir içerik besleme uygulamasıdır. XML formatında olan RSS dosyaları ilk Netscape firması tarafından geliştirilmiştir. RSS, internet sistemlerindeki güncellemelere kullanıcıların kolaylıkla erişip, izleyebilmeleri için site sahiplerinin kullandığı bir teknolojidir (http://www.emo.org.tr/ekler/aeac0a1c44a6a34_ek.pdf?dergi=).

2.3.2. Veri Ön İşleme

Veri ön işleme aşamasında eğitim ve sınıflandırma sürecinde kullanılacak olan başlangıç dökümanın hazırlığı yapılır. Bu aşama dokümandaki üzerindeki veride oluşabilecek problemleri çözmek ve bu verinin doğal yapısını çözerek daha anlamlı bilgiler üretmek ve kaliteli analiz yapabilmek için yapılır. Doküman bu aşamadan sonra bir diziye aktarılır. Veri ön işleme adımları Şekil 9'da gösterilmiştir.



Kaynak: Göker ve Tekdere,2017:295

Şekil 9. Veri ön işleme adımları

2.3.2.1. Dönüştürme

İnternette üzerindeki veriler genel de HTML veya XML formatlarında olduğundan bu veriler üzerinde herhangi bir işlem yapabilmek için veriler yapısal veri formatı olan düz metin haline dönüştürülmelidir. Yani içindeki belirli taglardan (
,<p> gibi) temizlenmelidir.

XML: Bir işaretleme dili olan XML ile veriler işaretlenir. İşaretlenmiş verilere eleman denir. XML belgeleri bazı kurallar ile düzenlenen elemanlardan oluşur.

XML'in avantajlarından biri de bu kuralların genişletilebilir olmasıdır.

- Yeni işaretler tanımlanabilir.
- Elemanların içerdiği verilerin doğru olup olmadığının kontrolü için bazı kurallar eklenebilir.
- Ortak bir kullanım için bu tanım ve kuralları mekanizmalar şeklinde tanımlayabiliriz.

Standart bir veri aktarım tekniği olan XML bankacılık başta olmak üzere tüm finansal işlemlerde kullanılır. XML, gelişmiş ERP yazılımlarının arasında ki haberleşmede de kullanılırlar. Çoğu sistem yazılımlarının ayar dosyaları XML formatındadır. Java Serileştirme mekanizması da nesnelere XML aracılığı ile kaydeder. XML dosyalarını kullanabilmemiz için belirli ayrıştırma yöntemleri vardır.

2.3.2.1.1. Ayrıştırma Yöntemleri

XML veya HTML formlarındaki verilerde işlem yapabilmemiz için veriyi uygun formata dönüştürmek gerekir bunu yapabilmek için de verilere ayrıştırma yapmalıyız. DOM ve SAX olmak üzere ayrıştırma işlemleri için iki değişik belge modeli sunar. Uygulama kapsamında DOM Parser yöntemi kullanılmıştır.

DOM: DOM (Document Object Model) dökümanı (belge) bir döküman ağacı olarak kabul eder. Bu ağacın her seviyesinde farklı DOM nesnelere olur. DOM ile ağaç üzerinde bulunan alt ağaçlar blok halinde işlenir. DOM doğrulanmış bir belge üzerinde belirli işlemleri yapma da veya karmaşık yapıda bulunan elemanları kullanmak için uygun olabilir.

SAX: SAX (Simple API for XML) dökümanı sırası ile okur. Örneğin ardışık bir şekilde gelen siparişlerin işlenmesinde SAX daha ideal olabilir. SAX hatalar ile karşılaştığı zaman hatalı elemanı atlayıp sıradaki elemana geçen uygulamalar için daha uygun olabilir.

2.3.2.2. Tarama ve İşaretleme

Bu aşama da metin içinde geçen simgeler noktalama işaretleri temizlenip anlam karmaşasına neden olmamak için tüm yapılar küçük harfe çevrilir.

2.3.2.3. Kök Bulma

Kelimelerin dökümünde ki frekansını doğru bir şekilde bulabilmek için bütün kelimeler kök haline getirilir. Kelime köklerinin bulunmasında Zemberek Kütüphanesinin kök bulma metodu kullanılmıştır.

Kelime	Kök
gör	gör
gördüm	gör
görmüştüm	gör
görmedim	gör
görsem	gör

Şekil 10. Aynı köke sahip kelimeler

Kök	Tip	Aldığı Ekler
düşün	Fiil	FIIL_GECMISZAMAN_DI
düşün	İsim	IMEK_HIKAYE_DI
düş	İsim	ISIM_TAMLAMA_IN, IMEK_HIKAYE_DI
düş	İsim	ISIM_SAHİPLİK_SEN_IN, IMEK_HIKAYE_DI

Şekil 11. Düşündü kelimesin kök ve eklerine ayrıştırılması

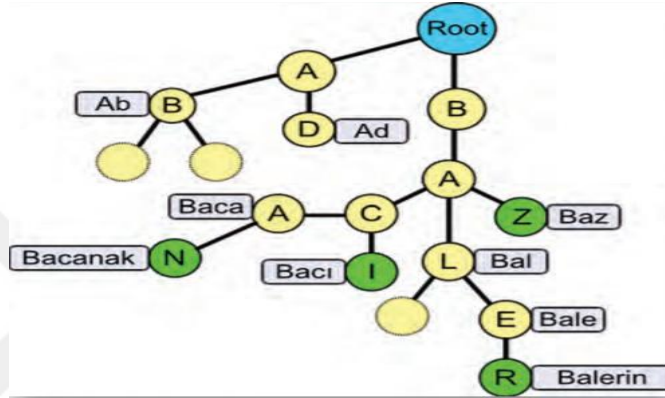
2.3.2.3.1. Cümle Çözümlemesi

Türk dil yapısına göre cümleler isim grubu ve fiil grubu olarak ikiye ayrılmıştır. İsim grubu sadece özne veya sadece nesneden oluşmasının yanı sıra her ikisinin birlikte kullanılması şeklinde de oluşabilir. Fiil grubu da sadece bir fiil veya fill ile beraber bir isim grubunun birlikte kullanımı şeklinde oluşabilir. Asıl hedef her bir grubu sözcüğe kadar ayrıştırabilmektir.

2.3.2.3.1.1. Zemberek Kütüphanesi

DDİ, gün geçtikçe yazılım dünyasında ki en zor ve popüler konulardan biri haline gelmektedir. Bu konuda çok fazla çalışmalar yapılmış olsa dahi, uygulama haline gelebilecek türdeki çalışmalar oldukça kısıtlıdır. Bu konu hakkında yapılan akademik

çalışmaların çoğunda Türk Dili ve Türkçe kullanılmıştır. Zemberek kütüphanesi oluşturulana kadar bu alan da neredeyse kullanılan açık kaynak kodlu bir kütüphane yoktu. Zemberek kütüphanesi sadece Türkçe için değil diğer Türk dilleri için de kullanılabilir bir temel DDİ sistemi sunmayı amaçlamaktadır. Zemberek kütüphanesi ile yazım denetimi, kök ve eklere ayırma, sözcük türetme, sözcük düzeltme, sözcük tipi bulma, sözcük önerme, sadece Türkçe'ye ait özel karakterlerin ASCII karşılığının kullanılması gibi temel DDİ işlemlerini sunmaktadır.



Kaynak: <http://zembereknlp.blogspot.com.tr/2007/02/zemberek-nasl-alr-1szlk-ve-kk-aac.html>

Şekil 12. Zemberek kütüphanesine göre kelime yapısı

2.3.2.4. Durak Kelimelerin Çıkarılması

Metin içinde çok fazla kullanılan ama tek başlarına kullanıldıklarında herhangi bir anlamı olmayan edat, bağlaç veya zamir gibi (stopwords) kelime türleri çıkartılır.

2.3.3. İndeksleme

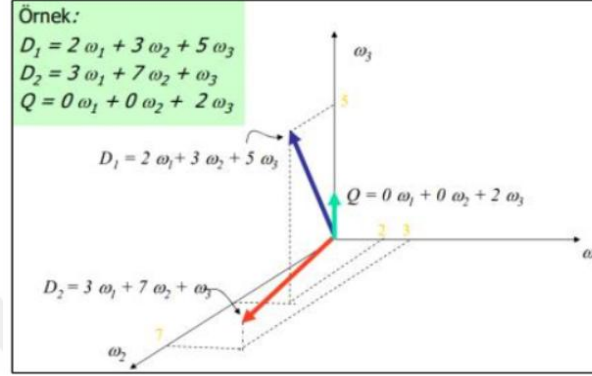
Doğal metinler içeren dökümanlar, sayısal bir veri türünde olmadıklarından dolayı bilgisayarlar tarafından tanınmazlar ve üzerlerinde herhangi bir işlem yapılamaz. Bundan dolayı ön işleme aşamasından geçen veri üzerinde işlem yapılabilmesi için dökümanın vektörel uzayda ifade edilmesi gerekir (Kaliyeva, 2013: 5).

2.3.3.1. Vektör Uzay Modeli

Matematiksel bir model olan vektör uzay modeli veri üzerinden bilgi çıkarımı, bilgi filtreleme ve indeksleme gibi işlemlerde kullanılır. Vektör uzay modeli, çok boyutlu uzayda doğal dil dökümanlarının özel bir anlamını simgelemektedir. Bir metnin

vektör ile ifade edilebilmesi için ilk olarak uzay eksenlerinin oluşturulması gerekir. Sınıflandırma sürecindeki sınıfın anahtar kelimeleri uzay eksenlerini oluşturur.

Dokümanlar Şekil 13’de gösterildiği gibi kelimelerin vektörleri olarak ifade edilir. (Di: eğitim dokümanları, Q:sorgu dokümanı Eksenler kelimeleri (Wi) ifade etmektedirler.



Kaynak: <http://www.csharpnedir.com/articles/read/?id=731>

Şekil 13. Vektör uzay modeli

Vektör uzay modelinde kelimeler ve dokümanlar i - j boyutlu bir uzayda gösterilir. Bu şekilde her doküman bir vektör olarak gösterilebilir. i 'nin değeri, incelenen kelime sayısıdır. j 'nin değeri ise incelenen doküman sayısıdır. Her bir boyut bir kelimeye karşılık gelmektedir.

Vektör uzay modelinin oluşturulmasında dokümanlardaki her bir kelimenin ağırlıklandırma önemlidir. Vektör uzay modeli ağırlıklandırma sonucu oluşturulur (Pilavcılar, 2007: 4).

2.3.3.2. Kelime Ağırlıklandırma

Kelimelerin kökleri elde edildikten sonra kelimelere ağırlık verme işlemi yapılır. Ağırlıklandırma işlemi kelimelerin o doküman için etkisinin ölçülmesi için yapılır (Karaca, 2012: 16). Ağırlıklandırma işleminde bit, sözcük frekansı, ters doküman frekansı ve sözcük frekansı-ters doküman frekansı kullanılır.

2.3.3.2.1. Bit Ağırlıklandırma

Kelimenin doküman içerisinde bulunup bulunmadığıyla ilgilenir. Eğer kelime doküman içerisinde geçiyorsa 1, geçmiyorsa 0 değeri verilir. Sözcüğün doküman da birden fazla geçip geçmesi önemsenmez sadece doküman da geçip geçmediğine bakılır (<http://www.csharpnedir.com/articles/read/?id=782>).

Döküman="Taraftarlar İstanbul'da bırakıldı."

Sözlük ="tarafatar", "İstanbul", "gol", "futbolcu"

Vektör =(1,1,0,0) şeklinde oluşur.

2.3.3.2.2. Sözcük Frekansı Ağırlıklandırma (TF)

Bir terimin doküman içinde sahip olduğu frekansdır. O kelimenin doküman da kaç kez geçtiğini gösterir (<https://bahadirakcora.wordpress.com/2016/03/13/vektor-uzayi-modeli-ve-tf-idf-agirliklandirmasi-2/>).

ft, d = terimin doküman içerisindeki adedidir.

tf(t,d) Genel olarak o kelimenin doküman içerisindeki geçme sayısıdır. Fakat duyarlılık ihtiyacına bağlı olarak farklı şekillerde de tanımlanabilir.

Tablo 4. Örnek Metinler

Id	Metinler
1	Gribe yakalanan hasta grip olduğunu anlamamıştı. İlacını almamıştı.
2	İlacını aksatanlar hastalığa davetiye çıkarırlar.
3	Yıllık enflasyon oranı bu senede yükselişte
4	Tarımla uğraşanlar bu yıl tarımdan zarar edecekler.
5	Hakemin gözü önünde olmasına rağmen hakem penaltı çalmadı. (Spor)
6	Taraftarlara erken gelen gol ilaç gibi geldi ve taraftarlar golden sonra hiç susmadı. (Spor)

Kaynak: İlhan vd., 2008:2

Sözlük={enflasyon, grip, hakem, ilaç, taraftar, tarım}

D1=(0,2,0,1,0,0)

D2=(0,0,0,1,0,0)

$$D3=(1,0,0,0,0,0)$$

$$D4=(0,0,0,0,0,2)$$

$$D5=(0,0,2,0,0,0)$$

$$D6=(0,0,0,1,2,0)$$

2.3.3.2.3. Ters Doküman Frekansı (IDF)

Bir terimin ne kadar eşsiz olduğunun hesaplanması işlemidir. IDF hesabı yapılırken “N” değeri toplam doküman sayısını gösterir. “df” ise hesabı yapılan kelimenin kaç dokümanda geçtiğini temsil eder. Cümledeki IDF değerinin yüksek olması analizi yapılan kelimenin kullanılmasının, kategorinin belirlenmesi için çok değerli olduğunu gösterir. Örneğin “çok” kelimesi bir cümlede tek başına bir duygu ifadesi içermemektedir. Bu nedenle hem olumlu cümle hem de olumsuz cümlenin içerisinde sık olarak geçebilir. TF-IDF yöntemi tam bu esnada bu kelimenin ağırlığını düşük bularak önemsiz olduğunun tespitinde büyük rol oynar.

$$TF = \frac{f}{df}$$

$$IDF = \log \left(\frac{N}{df} \right)$$

$$TF - IDF = TDF * IDF$$

2.3.4. Öznitelik Seçimi

Veri setlerinden öznitelik kümeleri elde etme işlemine öznitelik seçimi denir. Öznitelik seçimi ile verilerin hem doğru sınıflandırması hem de ölçeklenebilirliği açısından iyi bir sınıflandırma modeli elde edilir. Öznitelik seçme $M < N$ olmak üzere, N adet öznitelik içeren bir veri setinde ki belirli bir ölçüt fonksiyonu, M'nin tüm alt kümelerinde en iyi olacak şekilde M özniteliğinin belirlenmesi sürecidir.

Öznitelik alt kümesi oluşturulurken kullanılan stratejilere bağlı olarak iki temel yöntem vardır. Bunlar filtre tabanlı ve sarmalama tabanlı öznitelik seçim yöntemleridir.

Filtre tabanlı öznitelik seçim yönteminde belirlenmiş bir öznitelik altkümesinin yararlılığı, belirli bir sezgi ile değerlendirilmektedir. Bu yöntem, öznitelik

değerlendirmede kullandıkları yöntemlere dayalı olarak bireysel öznitelik ölçütleri ve grup öznitelik ölçütleri olarak iki sınıfa ayrılır. Bireysel öznitelik ölçütleri, özelliklerin uygunluk durumlarını bireysel olarak değerlendirir. Bireysel öznitelik değerlendirme ölçütlerinde en son oluşan öznitelik alt kümesi, tüm özniteliklerin bir ölçüte bağlı olarak sıralanması ve daha önceden belirlenen bir eşik değerinin üzerinde olan özniteliklerin seçilmesi ile oluşturulur. Grup öznitelik ölçütlerinde de özniteliklerin bireysel bir şekilde değerlendirilmez bunun yerine aday öznitelik altkümelerinin değerlendirilmesi yapılır. Böylece öznitelikler arası ilişkiler de işleme katılır.

Sarmalama tabanlı öznitelik seçim yönteminde de bir sınıflandırma algoritmasının başarısına bağlı olan öznitelikler seçilir. (Onan ve Korukoğlu, 2016: 2)

2.3.5. Sınıflandırma

Daha önceden belirlenmiş konu kategorilerinin yer aldığı bir belgede ki ana temaların tanımlanmasıdır. Metin sınıflandırma işlemi bu amacı gerçekleştirmek için kullanılan bir adım olup konuşma dili şeklinde yazılan metinleri anlamlarına göre önceden oluşturulmuş sınıflara ayırmaya çalışır. Metin sınıflandırma işlemi kontrollü bir kelime deposuna bağlı olarak belge indeksleme, belge filtreleme, otomatik metadatalar oluşturma ve web sayfalarını otomatik bir şekilde hiyerarşik bir sıra da düzenleme gibi pratik olarak uygulanabilen birçok alanda görmek mümkündür (Delen & Crossland, 2008; Tülek, 2007).

Sınıflandırma için pek çok yöntem bulunmaktadır. Bu yöntemler içinden en tercih edilen ve en önemlileri, karar ağaçları, yapay sinir ağları bulanık mantık, yaklaşımlı kümeler ve içerik öğrenmedir. Belge benzerliklerinin aranması da bir metin madenciliği uygulamasıdır ve aynı şekilde ön işleme, sınıflandırma ve kümeleme işlemlerini de içerir (Amasyalı, 2008).

2.3.5.1. Kümeleme

Kümeleme işlemi ile önceden belirlenmiş bir kategori kümeleri olmadan birbirlerine benzeyen verilerin gruplandırılması yapılır. Benzerlik tespitinde birden fazla teknik bulunmaktadır. Bunlar; Kosinüs benzerlik metodu, Jaccard benzerlik metodu ve Dice benzerlik metodu olup bir örnek üzerinden açıklanmıştır (<http://ab.org.tr/ab10/sunum/22.pdf>).

$$D = \{d_1, d_2, d_3, \dots, d_n\}$$

D öğrenci ödevleri olmak üzere d'ler her bir öğrencinin dokümanını temsil etsin. Değerlendirme sürecinde öncelikle d içinden metinler elde edilir.

$$d_i = \{w_1, w_2, w_3, \dots, w_n\}$$

di bir öğrencinin dosyası/dosyaları olmak üzere w gibi kelimelerden oluşur. İki öğrencinin ödevindeki kelimelerin karşılaştırılması için iç içe döngüye ihtiyaç vardır. Birinci döngüden elde edilen kelimeler ikinci döngüde elde edilen kelimelerle karşılaştırılır ve örtüşen kelime sayısı hesaplanır.

$$hd_i = \{(w_1, f_1), (w_2, f_2), (w_3, f_3), \dots, (w_n, f_n)\}$$

Burada w anahtar hash değeri ve f kelimenin geçme frekansı olmak üzere kelime havuzu aranır. Doküman havuzları hazırlandıktan sonra doküman benzerliklerinin bulunmasına geçilir.

Kosinüs benzerlik metodunda, iki vektör arasındaki kosinüs açısının ölçümünden elde edilen değer benzerlik için kullanılır. Bu değer aşağıdaki formül yardımıyla hesaplanmaktadır. Tez de benzer haber tespiti için kosinüs benzerlik metodu uygulanmıştır.

$$\cos_sim(d_x, d_y) = 1 - \frac{d_x \cdot d_y}{|d_x||d_y|} = 1 - \frac{x_i x_j + y_i y_j}{\sqrt{x_i^2 + x_j^2} \sqrt{y_i^2 + y_j^2}}$$

Burada, dX ve dY farklı iki doküman olmak üzere xij - yij benzer kelime sayısını ve xi- xj - yi- yj dokümanlardaki tüm kelime sayısını vermektedir. Jaccard Benzerlik Katsayısı, benzer eleman sayısının toplam kelime sayısına oranıdır.

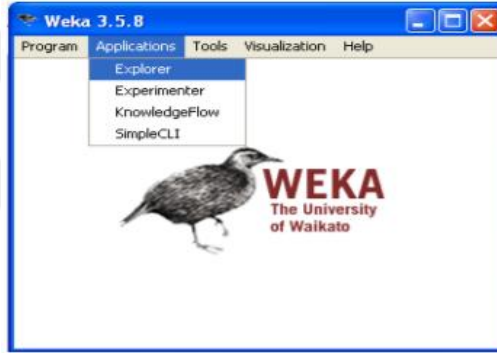
$$jac_sim(d_x, d_y) = 1 - \frac{d_x \cap d_y}{d_x \cup d_y} = 1 - \frac{x_i}{x_j + y_j - x_i}$$

Burada, xi - yi benzer kelime sayıları eşit olduğu için sadece xi alınmıştır. Başka bir deyişle xi iki dokümandaki benzer kelime sayısıdır. Dice katsayısı, “Lee Raymond Dice” tarafından farklı bir alan için önerilmiş fakat doküman benzerliğinde kullanılan bir ölçüttür (Uzun, Erdoğan ve Saygılı, 2016: 90).

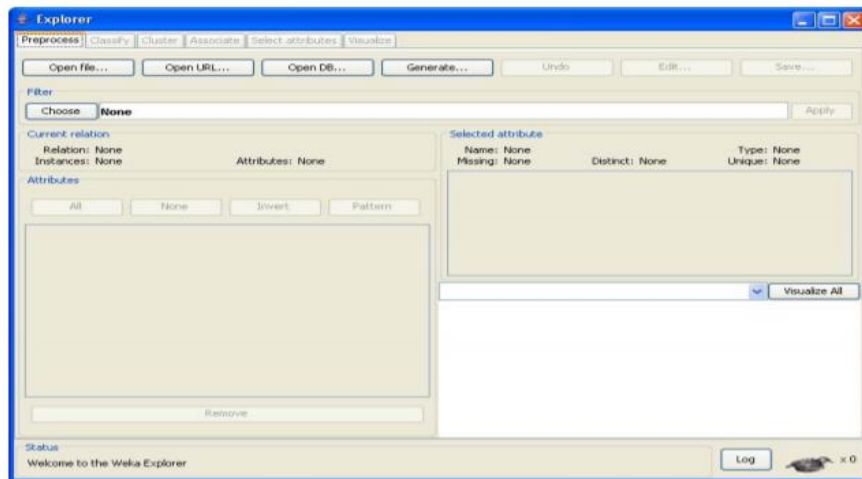
$$dice_{sim}(d_x, d_y) = 1 - \frac{2|d_x \cap d_y|}{|d_x| + |d_y|} = 1 - \frac{2x_i}{x_j + y_j}$$

2.4. Metin Madenciliği Aracı: WEKA

Bir proje olarak başlayan WEKA günümüzde birçok insan tarafından bir VM uygulaması geliştirme programı olarak kullanılmaktadır. Java platformunda geliştirilen WEKA, açık kaynak kodlu bir programdır. WEKA programının çalıştırılması ile Şekil 14’de görülen, Application menüsünde çalışılabilecek işlemler listelenmektedir. Bunlar komut modda çalıştırmaya izin veren Simple CLI, projeyi adım adım görsel bir içerikle gerçekleştirmeyi sağlayan Explorer ve projeyi sürükleyip bırak tekniği ile gerçekleştiren KnowledgeFlow seçenekleridir.



Şekil 14. WEKA’da Applications Menüsü



Şekil 15. WEKA’da Veri Kümesinin Seçimi

Metin (text) formatında ki verileri WEKA ile işlemek imkânsızdır. Bunun için verilerin WEKA programında çalışabilmesi için Arff veya Csv yapısında bulunan dosya formlarına dönüştürülmesi gereklidir. WEKA programı veri işleme, veri sınıflandırma, veri kümeleme, veri ilişkilendirme gibi özelliklere de sahiptir. Daha sonra projede yapılacak işleme uygun algoritma veya algoritmaların seçilip veriler üzerine uygulanması ile sonuçlar elde edilir. Tez de Arff dosya formatı kullanılmıştır.

Arff Dosyası: Arff dosyası bir metin dosyasıdır ve aşağıdaki yapıdan oluşmaktadır (https://web.uvic.ca/~maryam/DMSpring94/Labs/1_WekaIntro.pdf).

Arff dosyasının içeriği;

@relation Veri setinin ismi

@attribute Niteliğin tipi

@attribute Sınıflar {Sınıf1, sınıf2....}

@data

Değer1, değer 2, değer 3 , ...

2.4.1. Veri Eğitimi ve Test

Veriler seçildikten sonra gerekirse veri ön işleme aşamaları gerçekleştirilir. Ardından sınıflandırma aşamasına geçilir. Bu aşamada istenilen sınıflandırma algoritması seçilerek model eğitim kümesindeki veriler ile eğitilir. (Coşkun ve Baykal, 2011: 3- 4). Eğitim aşamasında dört seçenek kullanılabilir:

1. Eğitim kümesini kullanma (Use training set): Veri kümesi sadece modeli eğitmek için kullanılır.

2. Sınama kümesi ile sınama (Supplied test set): Veri kümesi “Use training set” seçeneğiyle eğitildikten sonra modelin doğruluğu ikinci bir küme ile sınanır.

3. Çapraz geçерleme (Cross-validation): Veri kümesi n eşit kümeye ayrılır ve eğitim için n-1 küme (validating set), sınama içinse 1 küme (test set) kullanılır. Bu işlem belirtilen küme sayısı kadar tekrarlanarak sistem iyileştirilmeye çalışılır.

4. Yüzde ayırma (Percentage split): Veri kümesini ikiye böler. Veri kümesinin belirli bir yüzdesi modelin eğitilmesi, kalan kısmı ise modelin sınanması için kullanılır (Şeker, 2013:54).

Bir modelin başarısını değerlendirme de kullanılan kavramlar hata oranı, kesinlik, duyarlılık ve F-ölçütüdür. Modelin başarısının ölçümü o modelde doğru

sınıflandırılan örnek sayısı ve yanlış sınıflandırılan örnek sayısının nicelikleriyle alakalıdır. Test sonucunda elde edilen sonuçların bilgileri karışıklık matrisi ile ifade edilebilir. Bu matris de satırlar test kümesinin gerçek sayıları, sütunlar ise modelin tahminlenmiş sonuçlarını ifade eder.

		Öngörülen Sınıf	
		Sınıf=1	Sınıf=0
Doğru Sınıf	Sınıf=1	a	b
	Sınıf=0	c	d

a: TP (True Pozitif) c: FP (False Pozitif)
b: FN (False Negatif) d: TN (True Negatif)

Şekil 16. Model Başarı Ölçütleri

2.4.1.1. Doğruluk- Hata Oranı

Modelin başarısının ölçümünde kullanılan en çok tercih edilen ve en kolay yöntem, modelin doğruluk oranıdır. Doğru olarak sınıflandırılmış örnek sayısının (TP +TN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır. Doğruluk oranını 1'e tamamlayan değere de hata oranı denir. Hata oranı yanlış sınıflandırılmış örnek sayısının (FP+FN), toplam örnek sayısına (TP+TN+FP+FN) oranı olarak ifade edilir.

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{Hata Oranı} = \frac{FP + FN}{TP + FP + FN + TN}$$

2.4.1.2. Kesinlik

Sınıfı 1 olarak tahmin edilen Doğru Pozitif örnek sayısının, sınıfı 1 olarak tahmin edilen tüm örnek sayısına oranıdır.

$$\text{Kesinlik} = \frac{TP}{TP + FP}$$

2.4.1.3. Duyarlılık

Dođru sınıflandırılmıř pozitif örnek sayısının toplam pozitif örnek sayısına oranıdır.

$$Duyarlılık = \frac{TP}{TP + FN}$$

2.4.1.4. F- Ölçütü

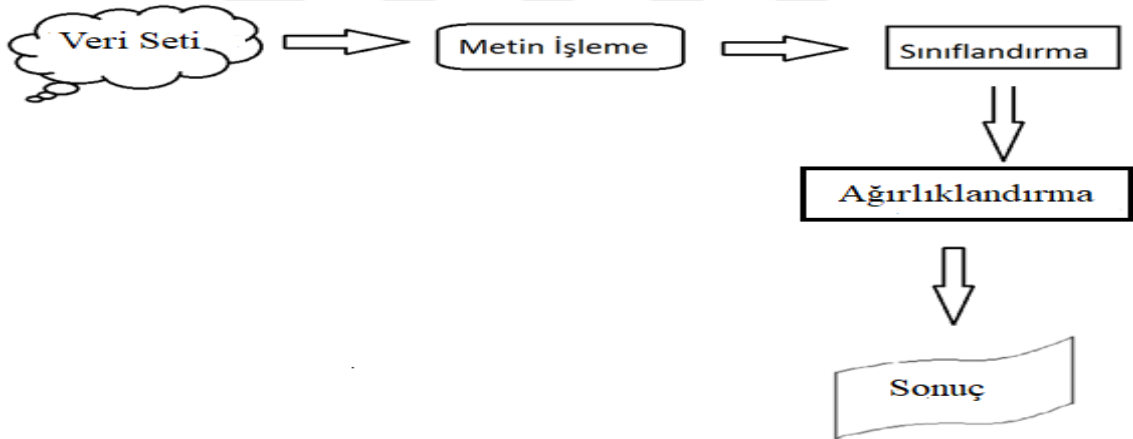
Kesinlik ve duyarlılık deđerleri ile anlamlı bir karşılaştırma sonucu çıkarmamız mümkün deđildir. Bu iki ölçütün birlikte kullanılmasıyla daha dođru sonuçlar elde edilecektir. İki ölçütün beraber kullanımını sađlamak için de f-ölçütü tanımlanmıştır. F-ölçütü, kesinlik ve duyarlılığın harmonik ortalamasıdır.

$$F - \text{Ölçütü} = \frac{2 \times \text{Duyarlilik} \times \text{Kesinlik}}{\text{Duyarlilik} + \text{Kesinlik}}$$

ÜÇÜNCÜ BÖLÜM

3. MATERYAL ve YÖNTEM

Bu bölümde tez de kullanılan materyal ve yöntemler ile tez çalışmasının sürecinden bahsedilmiştir. Tez çalışmasına ait adımlar Şekil 17’de gösterilmektedir. İlk adım olarak uygulama da kullanılacak verileri oluşturmaktır. Verileri toplamak için Dünya Gazatesinden haberler ile Borsa İstanbul resmi internet sitesi üzerinden belirli tarihler aralığında BİST100 endeksine ait veriler alınmıştır. Veri setleri hazırlandıktan sonra metin işleme yöntemleri uygulanarak veriler yapısal verilere dönüştürülmesi ile veriler öznitelik vektörleri halinde ifade edilmişlerdir. Öznitelik vektörleri şeklindeki veriler üzerine metin madenciliği yöntemleri uygulanıp eğitim ve test kümelerini oluşturarak sınıflandırma işlemine tabi tutulmuştur. Son adım da ise ağırlıklandırma yöntemleri ile haberler ağırlıklandırılarak etkilerinin ne kadar olduğu değerlendirilmiştir.



Şekil 17. Uygulamanın Genel Yapısı

Tez de gösterilen işlemleri gerçekleştirebilmek için Java programlama dili kullanılmıştır. Öznitelik ve sınıflandırma gibi işlemler için de WEKA Veri madenciliği uygulaması kullanılmıştır. RSS uygulaması ile alınan haber metinleri Java DOM Parser yöntemi ile uygun veri formatına dönüştürülmüştür.

Dünya Gazatesi, Türkiye'nin ilk ve tek günlük ulusal ekonomi gazetesidir. Makro ve mikro ölçekte ki haber ve analizleri yoluyla ekonomi gündemini takip eden, yorumlayan, iş ve finans dünyasının nabzını tutarak kamuoyu oluşmasına katkı sunar. Tez de Dünya Gazatesi resmi internet sitesinde 01.04.2016 ve 01.10.2016 tarihlerinde

yayınlanan ekonomi haberler RSS yöntemi ile elde edilmiştir. Şekil 19’da Dünya Gazetesinin RSS uygulaması gösterilmiştir.

3.1. Veri Setinin Oluşturulması

Bu çalışmanın amacı ekonomi haberlerinin BİST 100 endeksinin üzerine etkilerini araştırmaktır. Bunun için iki çeşit veri seti kullanılmaktadır. İlk veri seti ekonomik haber metinleri ikinci veri seti ise BİST 100 endeksine ait değerlerdir.

İlk veri seti olan haber metinleri için Dünya Gazetesinin internet sitesi üzerinden ekonomi başlığı altında yayınlanan haberler kullanılmıştır. Bu haber kaynağı ile 01.04.2016 ile 01.10.2016 tarihleri arasında yayınlanan 2328 haber toplanmıştır. Dünya Gazetesinin RSS uygulaması ile alınan haberlerin yapısı Şekil 19’da gösterilmektedir. Haberler incelenirken sadece başlık bölümleri taranmıştır. Haber veri setinden etkisizi kelimeler çıkarıldıktan sonra toplam 12453 farklı kelime bulunmuştur. Kelimeler arasında büyük küçük harften dolayı kelimeleri farklı olarak algılamaması için Java diline ait LowerCase metodu kullanılmıştır. Bunun sonucunda tüm kelimeler küçük harfe dönüştürülmüştür. Veri setini oluşturan dökümanın genel yapısı şu şekildedir.

Başlık	Ortadoğulu fonlar öğrenci evleri geliyor
İçerik	Yabancı fonlar, AVMlerde uzayan geri dönüşler nedeniyle rotayı öğrenci yurtlarına kaydirdi. Üniversitelerde kalan öğrencilerin sadece yüzde 10unun yurtlarda kalması öğrenci evlerine olan yatırım talebini arttıracakı belirtiliyor.
Yayınlanma Tarihi	01.04.2016

Şekil 18. Haberlerin genel yapısı

```

www.dunya.com/rss
is XML file does not appear to have any style information associated with it. The document tree is shown below.
<rss version="2.0">
  <channel>
    <title>Dünya Gazetesi</title>
    <link>http://www.dunya.com/link</link>
    <description>Dünya Gazetesi rss servisi</description>
  </channel>
  <item>
    <title>Kerkük'te gazeteci öldürüldü</title>
    <description>Kerkük'te yayın yapan radyo müdürü Muhammet Ubeydi, kimliği belirsiz kişilerin düzenlediği silahlı saldırı sonucu hayatını kaybetti.
    </description>
    <guid>340750</guid>
    <pubDate>Tue, 06 Dec 2016 22:51:00 +0300</pubDate>
  </item>
  <item>
    <title>Kiev'de ilk gol geldi</title>
    <description>UEFA Şampiyonlar Ligi'nde Besiktas, gruptaki son maçında evinde 1-1 berabere kaldığı Dinamo Kiev ile karşı karşıya geliyor.
    </description>
    <guid>340747</guid>
    <pubDate>Tue, 06 Dec 2016 21:58:00 +0300</pubDate>
  </item>

```

Şekil 19. Dünya gazetesinin RSS uygulaması

RSS üzerinden alınan verilerin XML dosya formatı Şekil 20'de gösterilmiştir.

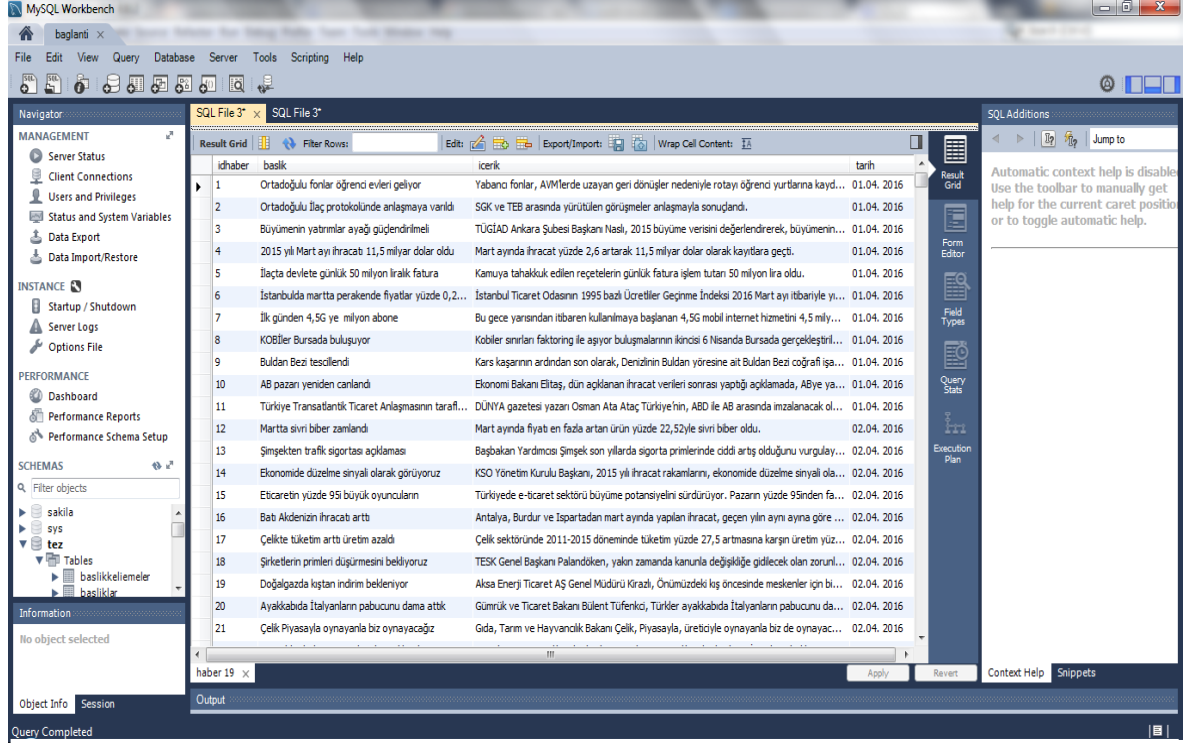
```

ekonomi - Not Defteri
Dosya Düzen Biçim Görünüm Yardım
<?xml-stylesheet type="text/xsl" href="/content/rss.xsl"?>
<rss version="2.0">
  <item>
    <title>Enerji ve Tabii Kaynaklar Bakanlığına önemli atamalar</title>
    <description>Enerji ve Tabii Kaynaklar Bakanlığına atamalar yapıldı. Yapılan atamalar Resmi Gazetede yayımlandı. </description>
    <pubDate>Wed, 22 Jun 2016 14:00:58 GMT</pubDate>
  </item>
  <item>
    <title>Coca-Cola, Uludağ Limonataya talip</title>
    <description>Coca-Cola, Uludağ markası ile gizli bir pazarlığa oturmuş. </description>
    <pubDate>Wed, 22 Jun 2016 10:58:38 GMT</pubDate>
  </item>
  <item>
    <title>Ruslardan ambargo itirafı</title>
    <description>Rusyada hazırlanan bir rapora göre, Türkiyeye yönelik gıda yaptırımları, rublenin devalüasyonu nedeniyle etkili olmadı </description>
    <pubDate>Wed, 22 Jun 2016 10:39:31 GMT</pubDate>
  </item>
  <item>
    <title>Rusya ile yaşanan uçak krizi çamur banyosunu da vurdu</title>
    <description>Rusya ile yaşanan uçak krizi sağlık turizmini de olumsuz etkiledi. Muğlanın Dalaman ilçesinde, daha çok Rus turistlerin tercih ettiği şifalı termal su ve çamur banyoları </description>
    <pubDate>Wed, 22 Jun 2016 08:16:07 GMT</pubDate>
  </item>
  <item>
    <title>Taşeron kadroda yaş sınırı</title>
    <description>Maliye Bakanlığının taşeron işçinin kadroya alınmasıyla ilgili yaptığı çalışmada yaş gelenler emekliliğe ayrılacak. Askerdeyken yerine personel alınan ve 1 Kasım 2015 </description>
    <pubDate>Wed, 22 Jun 2016 07:34:53 GMT</pubDate>
  </item>
  <item>
    <title>Memurlarla ilgili devrim niteliğinde karar</title>
    <description>Maaş Kütüğü adı verilen sistem ile milyonlarca memurun tüm maaş ve özlük haklarına ilişkin bilgiler tek çatı altında toplanıyor </description>
  </item>

```

Şekil 20. RSS ile ulaşılan haberlerin XML dosyası

XML formatındaki veriler üzerine gerekli ayrıştırıcı yöntemleri uygulanıp veriler haber formatı şekline dönüştürülerek veriler veritabanına kaydedilmiştir.



idhaber	baslik	icerik	tarih
1	Ortadođulu fonlar öğrendi evleri geliyor	Yabancı fonlar, AVM'lerde uzayan geri dönüşler nedeniyle rotayı öğrendi yurtlarına kayd...	01.04. 2016
2	Ortadođulu ilaç protokolünde anlaşmaya varıldı	SGK ve TEB arasında yürütülen görüşmeler anlaşmayla sonuçlandı.	01.04. 2016
3	Büyümenin yatırımlar ayağı güçlendirilmeli	TÜGİAD Ankara Şubesi Başkanı Naslı, 2015 büyüme verisini değerlendirek, büyümenin...	01.04. 2016
4	2015 yılı Mart ayı ihracatı 11,5 milyar dolar oldu	Mart ayında ihracat yüzde 2,6 artarak 11,5 milyar dolar olarak kayıtlara geçti.	01.04. 2016
5	İlaçta devlete günlük 50 milyon liralık fatura	Kanuya tahakkuk edilen reçetelerin günlük fatura işlem tutan 50 milyon lira oldu.	01.04. 2016
6	İstanbul'da martta perakende fiyatlar yüzde 0,2...	İstanbul Ticaret Odasının 1995 bazlı Ürettiler Geçirne İndeksi 2016 Mart ayı itibarıyla y...	01.04. 2016
7	İlk günden 4,5G ye milyon abone	Bu gece yarınsından itibaren kullanıma başlanan 4,5G mobil internet hizmetini 4,5 mily...	01.04. 2016
8	KOBİler Bursada buluşuyor	Kobiler sınırları faktoring ile ayıyor buluşmalarının ikincisi 6 Nisanda Bursada gerçekleştiril...	01.04. 2016
9	Buldan Bezci tesellendi	Kars kaşarının ardından son olarak, Denizlinin Buldan yöresine ait Buldan Bezci coğrafi işa...	01.04. 2016
10	AB pazarı yeniden canlandı	Ekonomi Bakanı Elitaş, dün açıklanan ihracat verileri sonrası yaptığı açıklamada, AB'ye ya...	01.04. 2016
11	Türkiye Transatlantik Ticaret Anlaşmasının taraflı...	DÜNYA gazetesi yazarı Osman Ata Ataç Türkiye'nin, ABD ile AB arasında imzalanacak ol...	01.04. 2016
12	Martta sivri biber zamlandı	Mart ayında fiyatı en fazla artan ürün yüzde 22,52'yle sivri biber oldu.	02.04. 2016
13	Şimşekten trafik sigortası açıklaması	Başbakan Yardımcısı Şimşek son yıllarda sigorta primlerinde ciddi artış olduğunu vurgulay...	02.04. 2016
14	Ekonomide düzelve sinyali olarak görüyoruz	KSO Yönetim Kurulu Başkanı, 2015 yılı ihracat rakamlarını, ekonomide düzelve sinyali ola...	02.04. 2016
15	Eticaretin yüzde 95 büyük oyuncuların	Türkiyede e-ticaret sektörü büyüme potansiyelini sürdürüyor. Pazarın yüzde 95'inden fa...	02.04. 2016
16	Batı Akdeniz'in ihracatı arttı	Antalya, Burdur ve Ispartadan mart ayında yapılan ihracat, geçen yılın aynı aynaya göre...	02.04. 2016
17	Çelikte tüketim arttı üretim azaldı	Çelik sektöründe 2011-2015 döneminde tüketim yüzde 27,5 artmasına karşın üretim yüz...	02.04. 2016
18	Şirketlerin primleri düşmesini bekliyoruz	TESK Genel Başkanı Palandöken, yakın zamanda kanunla değişikliğe gidilecek olan zorunl...	02.04. 2016
19	Doğalgazda kıstas indirim bekleniyor	Aksa Enerji Ticaret AŞ Genel Müdürü Kırazi, Önümüzdeki kış öncesinde meskenler için bi...	02.04. 2016
20	Ayakkabıda İtalyanların pabucunu dama attık	Gümrük ve Ticaret Bakanı Bülent Tüfenkci, Türkler ayakkabıda İtalyanların pabucunu da...	02.04. 2016
21	Çelik Piyasayla oynamanla biz oynayacağız	Gıda, Tarım ve Hayvancılık Bakanı Çelik, Piyasayla, üreticilerle oynamanla biz de oynayac...	02.04. 2016

Şekil 21. Haber dökümanının veritabanında gösterimi

İkinci veri seti olan BİST 100 endeksine ait veriler de Borsa İstanbul resmi internet sitesi üzerinden 01.04.2016 ile 01.10.2016 tarihleri arasında BİST 100 endeksine ait açılış fiyatı, fiyat (gün sonu kapanış fiyatı), fark, en yüksek fiyat, en düşük fiyat ve tarih değerleri alınmıştır. Bist 100 endeksine ait verilerin veritabanında gösterimi Şekil 22'de gösterilmiştir.

	idgerçekpiyasa	fiyat	fark	acilis	tarikh	enyukse	endusuk
2		71738.43	+ 0.20	72728.07	22.07.2016	72801.93	70426.16
3		71594.98	- 4.42	74902.84	21.07.2016	74902.84	71569.37
4		74902.84	- 1.67	76500.98	20.07.2016	76784.46	74597.25
5		76177.82	- 1.01	76220.41	19.07.2016	78338.21	75849.84
6		76957.61	- 7.08	80755.89	18.07.2016	81110.29	75427.59
7		82825.36	+ 0.29	82385.58	15.07.2016	82825.36	82234.14
8		82589.76	+ 1.56	81770.18	14.07.2016	82699.03	81591.07
9		81321.74	+ 0.22	81216.61	13.07.2016	81639.14	80917.20
10		81141.57	+ 2.41	79366.26	12.07.2016	81141.57	79250.09
11		79234.95	+ 1.56	78626.07	11.07.2016	79290.59	78626.07
12		78019.48	-0.47	78185.44	08.07.2016	78335.47	77667.86
13		78387.70	+ 0.56	78120.41	04.07.2016	78387.70	78000.11
14		77952.11	+ 1.48	77177.59	01.07.2016	77952.11	77177.59
15		76817.19	+ 0.14	76916.56	30.06.2016	77369.32	76498.71
16		76711.97	-0.28	76744.17	29.06.2016	77386.33	76355.14

Şekil 22. BİST 100 endeksine ait verilerin veritabanında gösterimi

3.2. Metin İşleme

Veritabanında kayıtlı olan haberler üzerinde gerekli veri madenciliği veya metin madenciliği yöntemlerini uygulayabilmek için haberleri yapısal veri formuna dönüştürmeliyiz. Dönüştürme işlemi için gerekli adımlar şunlardır:

- Metin dökümanında ki XML formatı ile gelen taglardan ve noktalama işaretleri ile sayısal ifadelerden temizlenmiştir. Şekil 20 incelendiği zaman XML formattaki veri description, pubdate, title gibi tagları içermektedir. Bu taglardan başlık ve diğer verileri alabilmek için XML yapısı üzerinde DOM Parser yöntemi kullanılarak veri taglardan ayrıştırılmıştır. Ayrıştırılan veri üzerinde ki noktalama işaretlerinden veya kelimeler arasında birden fazla boşluk durumlarından kurtulmak için Java da bulunan replaceALL metodu kullanılmıştır.

- Haber metinlerinde etkisiz kelime olarak kabul edilen (stop words) kelimelerin temizlenmesi. Bu kelime listesinin hazırlanmasında Türkçe için oluşturulan etkisiz kelime listesi kullanılmıştır. Tüm kelimeleri içeren bir liste oluşturulmuştur. Etkisi kelime listesi ile kelime listesi karşılıklı olarak taranıp kelime listesinde bu etkisiz kelimeler geçiyorsa silme işlemi yapılmıştır.

- Dokümanda ki düzenleme işlemlerinden sonra kelimeler kök halinde ifade edilmelidirler. Bunun için de Türk Dili için geliştirilen Zemberek Kütüphanesi kullanılmıştır. Zemberek kütüphanesinde bulunan kok metodu üzerinden kelimeCozumle işleminin çalıştırılması ile kelimelerin kökü bulunur. Bununla birlikte aynı köke sahip ama aldığı ekler ile farklı bir kelime gibi görünen kelimeler kök haline

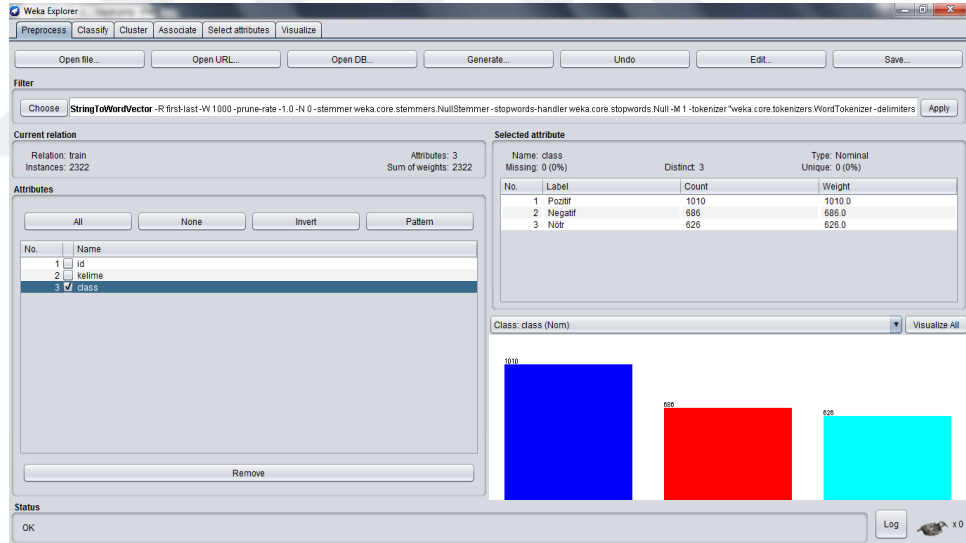
dönüştürülerek kelime uzayının boyutu indirgenmiştir. Örneğin gözlük, gözlükçü, gözlükçülük, göz kelimelere aynı köke sahip olsa da aldığı ekler sonucu farklı olarak görülmektedir. Bu tür oluşabilecek sorunları önleyebilmek için kelimeler kök hallerine dönüştürülmüşlerdir.

- WEKA uygulaması tf – idf vektör tanımlama ile öznitelik vektörleri oluşturulmuştur. Daha sonra bu kelimelerin geçme frekansları hesaplanmıştır. Hesaplanan frekans sayıları daha sonra kelimelerin ağırlıklandırmasında kullanılacaktır.

3.3. Sınıflandırma

Sınıflandırma işleminin gerçekleştirilebilmesi için öncelikli olarak bir eğitim ve test verisi bulunmalıdır.

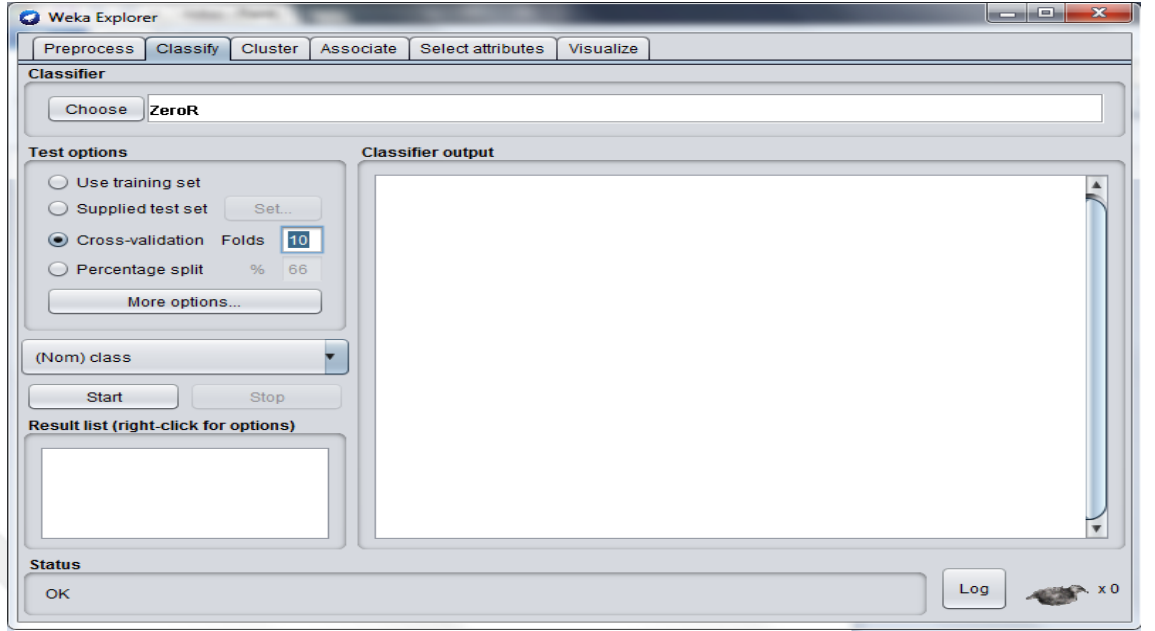
Veri kümesi 3 sınıf (pozitif/ negatif/ nötr) bilgisini içeren toplam da 2322 kelime kullanılıp bunların 1010 adeti pozitif, 686 adeti negatif ve 626 tanesi de nötr veri kaydından oluşmaktadır. Şekil 23’de özniteliklerin sınıflandırılması gösterilmiştir.



Şekil 23. Özniteliklerin Sınıflandırılması

Sınıflandırma işleminde WEKA yazılımı kullanıldığı için eğitim verisi k katlamalı çaprazlama tekniği ile veri seti eğitim ve test kümeleri olarak ayrılmıştır. Ayırma işleminde k değeri 10 olarak seçilmiştir. Bunun anlamı veri kümesi 10 eşit parçaya bölünmüştür ve bu 10 parçadan herhangi biri test için geri kalan 9 parça da eğitim için kullanılmıştır. Bu işlem her küme eğitim ve test işlemi oluncaya kadar devam etmiştir. En son sonuç kümesi de her adımda oluşan sonucun ortalaması alınarak

bulunur. K çaprazlama tekniğinin WEKA yazılımında uygulanması aşağı da gösterilmiştir.



Şekil 24. K çaprazlama tekniğinin uygulanması

Veri setleri oluşturulduktan sonra makine öğrenme algoritmalarından Naive Bayes, K-En yakın komşu ve temelinde destek vektör algoritmasını kullanan SMO kullanarak oluşturulan modellerin başarımları ölçütleri karşılaştırılmıştır:

Tablo 5. Sınıflandırma Algoritmalarının Karşılaştırılması

Değerler	Sınıflandırma Algoritmaları		
	SMO	Naive Bayes	k-NN (IBk k=1)
Doğru olarak sınıflandırılan örnek sayısı	1353	1123	1173
Yanlış olarak sınıflandırılan örnek sayısı	969	1199	1149
Doğru Pozitif (TP)	1015	994	924
Yanlış Negatif (FN)	716	673	772
Yanlış Pozitif (FP)	253	526	377
Doğru Negatif (TN)	338	129	249

Başarı Yüzdesi (Doğruluk)	58.2687 %	48.3635 %	50.5168 %
------------------------------	-----------	-----------	-----------

Tablo incelendiği zaman, başarı yüzdesi (doğruluk) en fazla olan algoritmanın SMO algoritması olduğu görülmektedir. SMO algoritmasının başarılı olmasının nedeni sınıf sayısının az olmasıdır.

En yüksek başarıyı gösteren SMO algoritmasının karışıklık matrisi (TP, TN, FP ve FN değerleri) incelendiğinde doğru pozitif sınıflandırılan örnek sayısı 1015, yanlış pozitif sınıflandırılan örnek sayısı 253, doğru negatif sınıflandırılan örnek sayısı 338 ve yanlış negatif sınıflandırılan örnek sayısı 716 Toplamda doğru sınıflandırılan örnek sayısı 1353 ve yanlış sınıflandırılan örnek sayısı 969 olarak tespit edilmiştir.

K- NN algoritmasının karışıklık matrisi (TP, TN, FP ve FN değerleri) incelendiğinde doğru pozitif sınıflandırılan örnek sayısı 924, yanlış pozitif sınıflandırılan örnek sayısı 377, doğru negatif sınıflandırılan örnek sayısı 249 ve yanlış negatif sınıflandırılan örnek sayısı 772 Toplamda doğru sınıflandırılan örnek sayısı 1173 ve yanlış sınıflandırılan örnek sayısı 1149 olarak tespit edilmiştir.

Navies Bayes algoritmasının karışıklık matrisi (TP, TN, FP ve FN değerleri) incelendiğinde doğru pozitif sınıflandırılan örnek sayısı 994, yanlış pozitif sınıflandırılan örnek sayısı 526, doğru negatif sınıflandırılan örnek sayısı 129 ve yanlış negatif sınıflandırılan örnek sayısı 673 Toplamda doğru sınıflandırılan örnek sayısı 1123 ve yanlış sınıflandırılan örnek sayısı 1199 olarak tespit edilmiştir.

3.4. Kelime Ağırlıklandırma

Haber dökümanlarına sınıflandırma algoritmaları uygulandıktan sonra ağırlıklandırma yöntemi uygulanarak oluşturulan öznitelik vektörlerinden etkili kelime listeleri oluşturulmuştur. Bu kelime listeleri 3 sınıftan oluşmaktadır. Bu listeler haber dökümanlarındaki haberlerin pozitif, nötr ve negatif etiketlerine göre oluşturulmuştur. Ağırlıklandırma işlemi de 2. bölümde anlatılan $Tf * Idf$ ile hesaplanmıştır. Bu kelimeler ve değerleri aşağıdaki tablolarda gösterilmiştir.

Tablo 6. Negatif Kelime Listesi

Kelime	Değer	Kelime	Değer
Belirsizlik	0.02981	Kayyum	0.09471
Boydak	0.03036	Kriz	0.12045
Brexit	0.1265	Negatif	0.03179
Ceza	0.241615	Olumsuz	0.091245
Dalgalanma	0.04144	Obama	0.04554
Darbe	0.66913	OHAL	0.145805
FETÖ	0.35475	Paralel	0.01518
Uyarı	0.26818	Risk	0.14322
Gözaltı	0.04554	Sığınmak	0.044055
Güneydoğu	0.0759	Soykırım	0.03366
İhraç	1.5444	Terör	0.0621
Temmuz	0.16082	Tartışma	0.05962

Tablo 7. Pozitif Kelime Listesi

Kelime	Değer	Kelime	Değer
Akkuyu	0.015895	Kazanmak	0.1012
Anlaşma	0.17688	Müjde	0.11748
Aselsan	0.09702	Olumlu	0.082
Avantaj	0.05698	Pozitif	0.03102
Babacan	0.03267	Reform	0.07722
Barış	0.1012	Sağlam	0.06318
Büyüme	0.43263	Tanap	0.04653
Elvan	0.10472	Güven	0.514745
İndirim	0.098615	Zeybekçi	0.08404

Tablo 8. Nötr Kelime Listesi

Kelime	Değer	Kelime	Değer
Abone	0.0759	Müzik	0.02018
Adres	0.01848	Mesai	0.02018
Kimlik	0.03509	İşbaşı	0.02018
Küba	0.017545	Anka	0.02018
Otobüs	0.044055	Yapay	0.01848
Trafik	0.182325	Mühür	0.01848
Yedek	0.03696	Oruç	0.01848
Torba	0.03179	Video	0.01848

Tablo 6 , Tablo 7 ve Tablo 8 incelendiği zaman BİST 100 endeksine olumlu yönde yani artışında neden olan kelimelere anlaşma, güven, büyüme vb. olmuştur. Olumsuz yönde yani azalmaya neden olan kelimelere ise ihraç, brexit, ceza, darbe, fetö gibi kelimelerdir. Nötr etkisiz olan kelimeler de değerlere sahiptir fakat bunların etkisiz olmaları bu kelimelerin geçtiği haberlerin nötr olarak sınıflandırılmasından dolayıdır.

İncelenen 2328 ekonomi haber dökümanından etkisiz kelimelerin (stop words) çıkarılması ve sadece haberin başlığı kullanıldığında 12453 kelimelik bir veri seti elde edilmiştir. Bu kelime uzayının boyutu büyük olmasından dolayı öznitelik seçimi işlemi yapılarak kelime uzayı 2322 kelimeye indirilmiştir. Taranan bu kelimeler incelendiği zaman etkisi 0.01 değeri altında olan kelimeler işleme alınmamıştır. Bu değer seçilme nedeni de bu etkiye sahip olan kelimelerin yer aldığı haberlerin yayımlandıkları güne ait borsa hareketleri incelendiği zaman borsa da ki değişim miktarında 0.001 değerinin altında olduğu gözlenmiştir.

SONUÇ ve ÖNERİLER

Yapılan bu tez çalışmasında haber sitesinde yayınlanan ekonomi haberlerinin BİST 100 endeksinin açılış ve kapanış fiyatları arasındaki fark değerine olan etkileri incelenmiştir. Toplam da 2328 adet ekonomi haberi toplanmıştır. Haber sitesinden alınan veriler işlenebilmek için uygun veri yapısına dönüştürülerek veri seti oluşturulmuştur. Elde edilen haberler cümleler halinde olduğundan bu yapılar kelime formuna dönüştürülüp veri seti oluşturulmuştur. Oluşturulan veri seti üzerinde metin işleme adımları uygulanıp veriler uygun veri yapısı haline getirilmiştir. Veri setinden ağırlıklandırma yöntemi ile öznitelikler elde edilmiştir. Bu öznitelikler durumlarına göre pozitif, negatif ve nötr olmak üzere 3 sınıfa ayrılmıştır. Metin işleme ve öznitelik vektörleri oluşturulduktan sonra sınıflandırma işlemi uygulanmıştır.

Sınıflandırma işleminin uygulanabilmesi için bir eğitim ve bir test verisi olmalıdır. Bu çalışma da eğitim ve test verisi , veri kümesi üzerinden K Kat Çaprazlama tekniği ile bulunmuştur. Bu teknikte literatür incelemesi yapıldığında genel de k değeri 10 olarak kabul edilmiştir ve bu tez çalışmasında da k değeri 10 olarak kabul edilip eğitim ve test verisi oluşturulmuştur. Eğitim ve test kümeleri oluşturulduktan sonra makine öğrenme algoritmalarından SMO, Navies Bayes ve KNN algoritmaları uygulanmıştır. Uygulama sonuçlarına göre en iyi sonuç SMO algoritması sonucu elde edilmiştir.

Öznitelik vektörlerine göre oluşturulan kelime listelerinin durumları da kelimelerin haber dökümanlarında sahip oldukları etikete göre atanmıştır. Kelime listeleri pozitif, negatif , nötr olarak 3 sınıfa ayrılmıştır. Daha sonra bu kelime listelerinin Tf* Idf hesaplamasına göre değerleri hesaplanmıştır.

Ağırlıklandırma işlemi sonucunda ihraç, darbe, fetö ve brexit kelimeleri en yüksek olumsuz değer olarak bulunmuştur. Bu kelimelerin geçtiği haber dökümanları incelendiğinde o gün ki BİST 100 endeksinin azaldığı görülmüştür. Büyüme, anlaşma, güven kelimeleri de olumlu değerler olarak bulunmuştur. Bu olumlu etkiye sahip kelimelerin geçtiği haberler incelendiği zaman ve o güne ait BİST 100 endeksinin arttığı görülmüştür. Nötr kelime listesindeki kelimelerinde değerleri hesaplanmıştır ancak bu kelimelerin geçtiği haber dökümanlarının etkisi olmadığı kabul edildiği için bu listedeki kelimeler de etkisiz olarak kabul edilmektedir.

İleri ki çalışmalarda haberler ekonomi haberleri ile sınırlı kalmayıp ulusal ve uluslararası siyasi, ekonomik ve diğer haberleri kapsayacak şekilde haber arşivi geliştirilirse ve birden fazla veri kaynağı kullanılıp, aralığı daha uzun olacak şekilde seçilirse her haberin veya her kelimenin nasıl etkilediğine dair daha detaylı sonuçlar elde edilecektir. Haber kaynağı seçimin de haberlere objektif olarak yaklaşan kaynaklar tercih edilirse veri tutarlılığı daha yüksek olacaktır. Kelime seçiminde tekli kelime grubu yerine ikişerli, üçerli vb. kelime grupları oluşturulursa haberlerin etkileri birden fazla değere sahip olacağından puanlama sisteminde daha başarılı sonuçlar elde edilir.



KAYNAKÇA

- Aghababaeayan, R. vd.** (2011), "Forecasting the Tehran Stock Market by Artificial Neural Network", International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence.
- Ahi, L.,** (2015) Veri Madenciliği Yöntemleri ile Ana Harcama Gruplarının Paylarının Tahmini, Hacettepe Üniversitesi, İstatistik Anabilim Dalı, Yüksek Lisans Tezi, Ankara.
- Akbulut, S.,** (2006) Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara.
- Akcan, A., C. Kartal** (2011) "İMKB Sigorta Endeksini Oluşturan Şirketlerin Hisse Senedi Fiyatlarının Yapay Sinir Ağları ile Tahmini", Muhasebe ve Finansman Dergisi, 07- 2011,27-40.
- Akkuş, İ.,** (2015) Veri Madenciliği - Kümeleme ve Kümeleme Yöntemleri.
- AÖF Dersleri** (2011), Menkul Kıymet Borsalarının Tarihçesi.
- Atasoy, Y.,** (2015) Veri Madenciliği Yöntemleri ile Ankilozan Spondilit Hastalığında Radyografik Progresyona Etkili Faktörlerin Analizi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Atiya A., Talaat N., ve Shaheen S.** (1997) An Efficient Stock Market Forecasting Model Using Neural Networks, Proc. Int. Conf. Neural Networks (ICNN), 2112-2115.
- Aykut,M.,** (2015) BİST-100 Endeksi ile Makroekonomik Değişkenler Arasındaki Nedensellik 2005-2015 Yılları Arasında Türkiye Uygulaması (Anadolu Üniversitesi Sosyal Bilimler Enstitüsü,Eskişehir.
- Bahadır, İ.,** (2008) Bayes Teoremi ve Yapay Sinir Ağları Modelleriyle Borsa Gelecek Değer Tahmini Uygulaması, TOBB Ekonomi ve Teknoloji Üniversitesi, Fen Bilimler Enstitüsü, Ankara.

- Birgöl, K. and Bertan, B.,** (2009). Yapay sinir ağları ile borsa endeksi tahmini. Yönetim, 20 (63): 25.
- Borsa İstanbul Uzmanları** (2014) “Finansal Piyasalar”, Sermaye Piyasası Lisanslama Sicil ve Eğitim Kuruluşu.
- Carvalho, A., T. Riberio** (2008), “Do Artificial Neural Networks Provide Better Forecasts, Evidence from Latin American Stock Indexes, Latin American Business Review, 8(3), 92-110.
- Ceran, G.,** (2006) Esnek Akış Tipi Çizelgeleme Problemlerinin Veri Madenciliği ve Genetik Algoritma Kullanılarak Çözülmesi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Konya.
- Chan, K.C.C., Teong, F.K.,** (1995) Enhancing Technical Analysis in the Forex Market Using Neural Networks, Neural Networks, Proceedings, IEEE International Conference, 2, 1023-1027.
- Çalış, A., Kayapınar, S. ve Çetinyokuş T.,** (2014) “Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama” Endüstri Mühendisliği Dergisi Makale Cilt: 25 Sayı: 3-4.
- Çalışkan, S. ve Soğukpınar, İ.,** (2008) “K&KNN: K- Means ve K En Yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti”.
- Demirel, H.G.,** (2009) Krizlerin Menkul Kıymetler Borsalarına Etkisi: İMKB Örneği, Karamanoğlu Mehmet Bey Üniversitesi Sosyal Bilimler Enstitüsü, Karaman.
- EMO,** İzmir Şubesi Bilgisayar MDK RSS Nedir?
- Ergün, K.** (2012) “Metin Madenciliği”.
- Erdoğan, E., H. Özyürek** (2012), “Yapay Sinir Ağlarıyla Fiyat Tahminlenmesi”, Sosyal ve Beşeri Bilimler Dergisi, 4(1), 1309-8012.
- Gidófalvi, G.** (2001). Using News Articles to Predict Stock Price Movements.

- Grudnitski, G. ve L. Osburn** (1993), "Forecasting S&P and Gold Future Prices: An Application of Neural Network", The Journal of Future Markets, 13(6), 631-643.
- Gülçe, G.**, (2010) Veri Ambarı ve Veri Madenciliği Teknikleri Kullanılarak Öğrenci Karar destek Sistemi Oluşturma, Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Edirne.
- Güngör, M.**, (2009) Türkiye’de Uygulanan Para Politikasının Sermaye Piyasası Karlılığına Etkisi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü ,İstanbul, s:78.
- Hotho, A., Nurnberger, A., Paaß, G.,A** (2005) Brief Survey of Text Mining. LDV Forum - GLDV Journal for Computational Linguistics and Language Technology 20(1), 19-62.
- Gündüz, H.**, (2013) BIST 100 Endeksi Yönünün Ekonomi Haberleri ile Tahmin Edilmesi, İstanbul Teknik Üniversitesi , Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Göker, H.**, (2012) Üniversite giriş sınavında öğrencilerin başarılarının veri madenciliği yöntemleri ile tahmin edilmesi, Gazi Üniversitesi, Bilişim Enstitüsü ,Yüksek Lisans Tezi, Ankara.
- İlhan, S., vd**, 2008. "Metin Madenciliği ile Soru Cevaplama Sistemi." 356-359.
- Jabin, S.** (2014), "Stock Market Prediction Using Feed-forward Artificial Neural Network", International Journal of Computer Applications, 99 (9).
- Kara, Y. ve diğ.** (2011), "Predicting Direction of Stock Price Index Movement Using Artificial Neural Networks and Support Vector Machines: The sample of The Istanbul Stock Exchange", Expert Systems with Applications 38, 5311–5319.
- Karakuş, K.**, (2009) Veri Madenciliği Teknikleri ile Mobil Telekom Sektöründe Müşterilerin Kredi Skorlamasına İlişkin İstatistiksel Bir Analiz, Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul.

- Kumdreli , Ü.C.,** (2012) Tıp Bilişimi ve Veri Madenciliği Uygulamaları: EEG Sinyallerindeki Epileptiform Aktiviteye Veri Madenciliği Yöntemlerinin Uygulanması, Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Edirne.
- Lam, M.,** (2004) Neural Network Techniques for Financial Performance Prediction: Integrating Fundamental and Technical Analysis, ELSEVIER, Decision Support Systems, 37, 567-581, 2004.
- Li, R., Xiong, Z.,** (2005) Forecasting Stock Market With Fuzzy Neural Networks, Machine Learning and Cybernetics, Proceedings of 2005 International Conference on, 6, 3475-3479.
- Liu Q., Lu X., Ren F., Kuroiwa S.,** (2004) Automatic Estimation of Stock Market Forecasting and Generating the Corresponding Natural Language Expression, International Conference on Information Technology: Coding and Computing (ITCC'04).
- Lunden, I.** (2014) Thomson Reuters Taps Into Twitter For Big Data Sentiment Analysis. TechCrunch. <http://social.techcrunch.com/2014/02/03/twitter-raises-its-enterprise-cred-with-thomson-reuters-sentiment-analysis-deal/>.
- Majhi, R., Panda, G., Sahoo, G.,** (2006) Efficient Prediction of Foreign Exchange Rate Using Nonlinear Single Layer Artificial Neural Model, Cybernetics and Intelligent Systems, 2006 IEEE Conference on, 1-5.
- McNelis, P.D.** (1996), "A Neural Network Analysis of Brazilian Stock Price: Tequila Effects vs. Pisco Sour Effects", Journal of Emerging Markets, 1(2).
- MEGEP** (Mesleki Eğitim ve Öğretim Sisteminin Güçlendirilmesi Projesi) (2007) Gazetecilik alanı haberin nitelikleri, Ankara.
- MEGEP** (Mesleki Eğitim ve Öğretim Sisteminin Güçlendirilmesi Projesi) (2007) Muhasebe Finansman Finansal İşlemler, Ankara.
- Mittermayer, M.,** (2004). Forecasting intraday stock price trends with text mining techniques, 37th Annual Hawaii International Conference on System Sciences, Proceedings of the, vol. 00, 64-73.

- Onan, A., ve Korukoğlu, S.,** (2016) Metin Sınıflandırmada Öznitelik Seçim Yöntemlerinin Değerlendirilmesi Akademik Bilişim.
- Özalp, A., A.S. Anagün** (2001), “Hisse Senedi Fiyat Tahmininde Yapay Sinir Ağı Yaklaşımı ve Klasik Tahminleme Yöntemleri ile Karşılaştırılması”, Endüstri Mühendisliği Dergisi, 12(3-4), 2-17.
- Özmen, Ş.,** (2001) "İş hayatı Veri Madenciliği ile İstatistik Uygulamalarını Yeniden Keşfediyor", V. Ulusal Ekonometri ve İstatistik Sempozyumu, Adana.
- Özyurt, M.O.,** “Borsa Nedir, Ne İçin Ortaya Çıkmıştır Ve İşleyişi Nasıldır? ” ,Gazete Bilkent, 31.03.2015
- Patel, M. B., S.R. Yalamalle,** (2014) “Stock Price Prediction Using Artificial Neural Network” International Journal of Innovative Research in Science, Engineering and Technology ,3(6).
- Pavlidis, N.G., Tasoulis, D.K., Vrahatis, M.N.,** (2003) Financial Forecasting Through Unsupervised Clustering and Evolutionary Trained Neural Networks, Evolutionary Computation, CEC '03. The 2003 Congress on, 4, 2314-2321.
- Xiong, Q., Yong, S., Shi, W., Chen, J., Liang, Y.,** (2005) The Research of Forecasting Model based on RBF Neural Network, Neural Networks and Brain, ICNN&B '05. International Conference on , 2,1032-1035.
- Schierholt, K., Dagli, C.H.,** (1996) Stock Market Prediction Using Different Neural Network Classification Architectures, Computational Intelligence for Financial Engineering, Proceedings of the IEEEIAFE 1996 Conference on, 72-78, 24-26.
- Schumaker, R. P., Zhang, Y., Huang, C.-N. ve Chen, H.** (2012). Evaluating sentiment in financial news articles. Decision Support Systems, 53(3), 458–464. doi:10.1016/j.dss.2012.03.001.
- Seker, S.E.,** (2015) Metin Madenciliği, YBS Ansiklopedisi, 2015.

- Shah, M. et al.** (2014), “Performance Analysis of Neural Network Algorithms on Stock Market Forecasting”, *International Journal Of Engineering And Computer Science*, 3(9), 8347-8351.
- Şahin, Ö.**, (2014) Bist'teki Endekslerin Volatilitelerinin Karşılaştırılmalı Analizi: Bist Kurumsal Yönetim, Bist 100, Bist 50, Bist 30 Endeksleri Üzerinde Bir Uygulama, Düzce Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi, Düzce.
- Şeker, Ş. E.**, (2013) “İş Zekası ve Veri Madenciliği”, Cinius, İstanbul, 54.
- Taşkın, Ç. ve Emel, G.**, (2010) “Veri Madenciliğinde Kümeleme Yaklaşımları ve Kohonen Ağları ile Perakendecilik Sektöründe Bir Uygulama”.
- Telli, Ş.**, (2016) BIST 100 Endeksinin Ekonomik Takvim Etkisi Gözetilerek Yapay Sinir Ağları ile Öngörülmesi, Anadolu Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Eskişehir.
- Ticknor, J.L.** (2013), “A Bayesian Regularized Artificial Neural Network for Stock Market Forecasting”, *Expert Systems with Applications*, 40, 5501-5506.
- Toraman, C.** (2008), “Demir Çelik Sektöründe Yapay Sinir Ağları ile Hisse Senedi Fiyat Tahmini: Erdemir A.Ş. ve Kardemir A.Ş. Üzerine Bir Tahmin Uygulaması”, *Muhasebe ve Finansman Dergisi*; 39, 20-32.
- TSPAKB**, Türkiye Sermaye Piyasası Aracı Kuruluşları Birliği (2012) Menkul Kıymetler ve Diğer Sermaye Piyasası Araçları.
- Uzun, E., Erdoğan, C. ve Saygılı A.**, (2016) Kümeleme Modeli Kullanan Web Tabanlı Bir Ödev Değerlendirme Sistemi, *Electronic Journal of Vocational Colleges*.
- Vaisla, K.S. ve A.K. Bhatt** (2010), “An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting”, *International Journal on Computer Science and Engineering*, 2(6), 2104-2109.
- Yao, J., Tan, C.L.**, (2000) Time Dependent Directional Profit Model for Financial Time Series Forecasting, *Neural Networks, IJCNN, Proceedings of the IEEEINNS-ENNS International Joint Conference on*, 5, 291-296.

Yalta, Y., (2011) Finansal Piyasalar ve Finansal Kurumlar

White, H. (1988), "Economic Prediction Using Neural Network: The Case of IBM Daily Stock Return", IEEE International Conference on Neural Networks, 2(1), 451-458.

Wong, C. C., Chan, M. C., Lam, C., (2000)"Financial Time Series Forecasting By Neural Network Using Conjugate Gradient Learning Algorithm And Multiple Linear Regression Weight Initialization" Paper provided by Society for Computational Economics in its series Computing in Economics and Finance

Wuthrich, B., Permuntilleke, D., Leung, S., Cho, V., Zhang, J. ve Lam, W. (1998). Daily prediction of major stock indices from textual www data, in KDD, 364–368.

Zhora, D. Y., (2005) Data Preprocessing for Stock Market Forecasting Using Random Subspace Classifier Network, Proceedings of the International Joint Conference on Neural Networks, Montreal., 2549-2554, 1005.

URL 1- <http://www.gazetebilkent.com/2015/03/31/borsa-nedir-ne-icin-ortaya-cikmistir-ve-isleyisi-nasildir/>

URL 2- <http://www.borsaistanbul.com>

URL 3- https://tr.wikipedia.org/wiki/Borsa_%C4%B0stanbul

URL 4- <http://bizdosyalar.nevsehir.edu.tr/251e26563cc3fddabe6ce826c045b91d/spl.pdf>

URL 5- <http://borsanasiloynanir.co/borsa-gundem-olaylarindan-nasil-etkilenir/>

URL 6- <http://borsanasiloynanir.co/borsa-tarihi/>

URL 7- <http://www.belgeci.com/dunya-borsasinin-tarihi.html>

URL 8- <http://www.endeksnedir.com/>

URL 9- <http://www.borsahocasi.net/2015/02/dunya-borsalari.html>

URL10- http://bankamufettisleri.blogspot.com.tr/2013/07/bist-ve-bist-endeksleri_nedir.html.

URL 11- <http://www.bireyselemeklilikbes.com/bist-50-endeksi-nedir/>

URL 12- <https://ahmetsedef.wordpress.com/tag/veri-madenciligi-sureci/>

URL 13- <http://ab.org.tr/ab13/bildiri/175>

URL 14- http://ybsansiklopedi.com/wpcontent/uploads/2015/08/MetinMadenciligi30_32.pdf

URL 15- <http://visualdatamining.blogspot.com.tr/2009/06/veri-madenciliginin-yararlar-ve.html>

URL 16- http://en.karsan.com.tr/images/yatirimci_iliskileri/s_sorulan_sorular/sorularlar_borsa_sermaye_piyasasi.pdf

URL 17- <http://www.reelpiyasalar.com/kunye/>

EKLER

Ek-1: Tez Orijinallik Raporu



SOSYAL BİLİMLER ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ÇALIŞMASI ORJİNALLİK RAPORU

ÖĞRENCİ BİLGİLERİ	
Adı-Soyadı	Özlem ALPAY
Öğrenci Numarası	141217109
Enstitü Anabilim Dalı	Teknoloji ve Bilgi Yönetimi
Programı	Teknoloji ve Bilgi Yönetimi
Danışmanın Unvanı, Adı-Soyadı	Yrd. Doç. Dr. Cem AYDEN
Tez Başlığı (Türkçe)	Ekonomi Haberlerinin BİST 100 Endeksine Etkisinin Veri Madenciliği ile İncelenmesi

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE

Yukarıda başlığı belirtilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam ...67.....sayfalık kısmına ilişkin, 18/09/2017 tarihinde Sosyal Bilimler Enstitüsü tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezinin benzerlik oranı % 21 'dir.

Uygulanan filtrelemeler:

- 1- Kabul/Onay ve Bildirim sayfaları hariç,
- 2- Kaynakça hariç
- 3- Alıntılar hariç/dâhil
- 4- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Yukarıda bilgileri verilen öğrencinin yüksek lisans tezi Sosyal Bilimler Enstitüsü Yönetim Kurulu tarafından belirlenen azami benzerlik oranlarını aşmadığını ve tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim. Cereğini saygılarımla arz ederim.

Yrd. Doç. Dr. Cem AYDEN
Danışmanın Adı-Soyadı
(İmzası)

Yrd. Doç. Dr. Cem AYDEN
Anabilim Dalı Başkanı
(İmzası)

F.Ü.LİSANSÜSTÜ EĞİTİM ÖĞRETİM YÖNETMELİĞİ

Madde 41- Lisansüstü tezleri ile birlikte teslim edilmesi gereken belgeler şunlardır:

- a) Lisansüstü tezler, savunma öncesinde intihal program raporu ve ilgili makale şartını sağladığına dair belgeleri ile birlikte enstitüye teslim edilir.
- b) İntihal raporu ile ilgili olarak etik kurallar dâhilindeki benzerlik oranları ilgili Enstitü Yönetim Kurulu tarafından belirlenir. (Enstitü Yönetim Kurulu tarafından tezin, intihal kapsamı dışında değerlendirilmesi için TURNITIN'den alınan raporda "benzerlik oranı"nın, "alıntılar hariç" en fazla %10, "alıntılar dâhil" % 30'u geçmemesi şeklinde kabul edilmiştir).

ÖZGEÇMİŞ

1992 yılında Elazığ’ da doğan Özlem ALPAY, orta ve lise öğrenimini sırasıyla Mezre İlköğretim Okulu ve Mehmet Koloğlu Anadolu Lisesinde tamamlamıştır. 2010 yılında kazandığı Fırat Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümünü 2014 yılında başarıyla bitirmiştir. 2014 yılında yüksek lisans eğitimine Fırat Üniversitesi Sosyal Bilimleri Enstitüsü Teknoloji ve Bilgi Yönetimi Anabilim Dalında başlamıştır. 2016 yılında yüksek lisans eğitimine Fırat Üniversitesi Fen Bilimleri Enstitüsü Donanım Anabilim Dalında başlamıştır. 2017 yılında Recep Tayyip Erdoğan Üniversitesi Bilişim Sistemleri Mühendisliği Bölümüne Araştırma Görevlisi olarak atanmıştır.

