

KADİR HAS ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ



EKONOMİ HABERLERİNİN BIST100 VE HİSSE SENETLERİNİN  
FİYAT DEĞİŞİMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

Engin Özer

Ocak, 2015

Engin Özer

Yüksek Lisans Tezi

2015 |

EKONOMİ HABERLERİNİN BIST100 VE HİSSE SENETLERİNİN  
FİYAT DEĞİŞİMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ



ENGİN ÖZER

Finans Mühendisliği Programı Yüksek Lisans derecesi  
için gerekli kısmi şartların yerine getirilmesi amacıyla  
Fen Bilimleri Enstitüsü'ne  
teslim edilmiştir.

KADİR HAS ÜNİVERSİTESİ

Ocak, 2015

KADIR HAS ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ



EKONOMİ HABERLERİNİN BIST100 VE HİSSE SENETLERİNİN  
FİYAT DEĞİŞİMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

ENGİN ÖZER

ONAYLAYAN:

Yard. Doç. Dr. Arif Selçuk Öğrenci (Kadir Has Üniversitesi  
(Tez Danışmanı)

Doç. Dr. Songül Albayrak (Yıldız Teknik Üniversitesi)

Doç. Dr. Mehmet Nafiz Aydın (Kadir Has Üniversitesi)

ONAY TARİHİ: 23/ 1 / 2015

“Ben Engin Özer, bu Yüksek Lisans tezinde sunulan çalışmamda başka çalışmalardan yaptığım alıntıların kaynaklarını kurallara uygun biçimde tez içerisinde belirttiğimi ve çalışmanın şahsıma ait olduğunu onaylıyorum.”

  
\_\_\_\_\_

# EKONOMİ HABERLERİNİN BIST100 VE HİSSE SENETLERİNİN FİYAT DEĞİŞİMLERİ ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

## Özet

Bu çalışmada ilk olarak, ekonomi haberlerinin belli bir kaynaktan toplanması, Bag-of-Words (ön tanımlı kelime grupları) yöntemi ile pozitif ve negatif olarak ayrıştırılması için bir yazılım geliştirilmiştir. Daha sonra bu ayrıştırma işleminin başarı oranı ele alınmıştır. İstanbul Borsası endeksindeki hareketlerin pozitif-negatif haberlerden ne kadar etkilendiği ve ileri yönelik tahmin yapılıp yapılamayacağı incelenerek bu gözlemlere dayalı bir portföy yönetim stratejisi önerilmiştir. Literatürde kullanılan algoritmalar kodlanarak ilgili ekonomi haberleri (siyasi, güncel, magazin vs. haber grupları çalışma dışında tutulmuştur) puanlanmıştır. Bu puanlamalar farklı algoritmalar ile karşılaştırılmıştır. Elde edilen sonuçlar üzerinden BIST100 ve 5 adet seçilmiş şirket için, ilgili saatlerdeki fiyat hareketlerinin haber puanlarıyla karşılaştırılması yapılmıştır. Fiyat hareketlerinde saatlik ve günlük fiyat değişimleri kullanılmıştır. Seans bitiminden sonra çıkan haberlerin seans açılışındaki etkisi ve yine seans içindeki yönün tahmini yapılmaya çalışılmıştır. 12 aylık ekonomi haber arşivi (51.878 adet) ve BIST100 fiyat geçmişi üzerinde çalışılmıştır.

Yapılan çalışmalarda, günlük ve saatlik periyottaki ekonomi haberlerinin fiyat hareketleri üzerine bir etkisi olduğuna dair ampirik bir kanıt bulunamamıştır. Bu nedenle haber puanlarını kullanarak ileri dönük bir tahmin yapılamaz. Öte yandan, gözleme dayalı olarak elde edilen bir desen dikkati çekmektedir: belirli süreli sürekli pozitif (veya negatif) haber akışının arkasından ani ve nisbeten yüksek değişimler gelebilmektedir. Bu desene dayanılarak geliştirilen portföy stratejisi ise yüksek getiri konusunda umut veren sonuçlar üretmiştir.

**Anahtar Kelimeler,** Bag-of-Words, Aho-Corasick, Türkiye Ekonomik Psikoloji Sözlüğü, Örümcek, Ekonomi Haberleri, Portföy Stratejisi, Maximum Entropy, Korelasyon

# ANALYSIS OF EFFECTS OF ECONOMICS NEWS ON THE PRICE FLUCTUATIONS IN BIST100 AND STOCKS PRICES

## **Abstract**

A software package has been developed in this thesis. The software collects economics news from a specified source automatically, and then it categorizes the news as positive or negative according to the Bag-of-Words method (using a predefined set of word groups). Then, the success rate of this classification process is investigated. It is examined how much fluctuations in the Istanbul Stock Exchange are affected by the degree of the news in order to investigate the possibility of a future prediction. Finally, a portfolio management strategy is proposed that is based on those observations. Related economics news (political, local, magazine etc. news have been excluded) are rated using the algorithms mentioned in the literature. Those ratings are compared using different algorithms. Based on those results, comparison of news ratings and price movements of BIST100 and 5 selected company stock prices, has been accomplished. Hourly and daily changes have been used in price movements. News appearing after the stock market closure have been investigated for their effects on market opening. Essentially, efforts have been spent to predict the direction of the price movements during the market session. The dataset is comprised of 51878 news spanning a period of 12 months, and the related BIST100 price history.

No empirical evidence is found in this study that the daily and hourly economics news have an effect on price fluctuations. Hence a future prediction based on the ratings of the news seems to be impossible. On the other hand, an observation based pattern requires attention: abrupt and relatively high fluctuations may occur following the flow of news that exhibit either positive or negative ratings for a specific period. The portfolio management strategy based on this pattern produced promising results for achieving high returns.

**Keywords:** Bag-of-Words, Aho-Corasick, Turkey Economics-Psychology Dictionary, Web Spider, Economics News, Maximum Entropy, Portfolio Strategy, Correlation

## İçindekiler

Özet.....	II
Abstract.....	III
İçindekiler.....	IV
Şekil Listesi.....	V
Tablo Listesi.....	VI
Kısaltmalar.....	VII
Giriş.....	1
1. Haberler.....	3
1.1. Haber Tipleri.....	4
1.2. Haberlerin Yatırım Araçları Üzerindeki Etkisi.....	5
1.3. Haberlerin İşlenmesi ve Değerlendirilmesi.....	6
2. Yöntem.....	9
2.1. Aho-Corasick Dizi Arama Algoritması.....	16
2.2. Bag-of-words Model Tabanlı Genişletilmiş Sorgulama.....	18
2.3. Türkiye Ekonomik Psikoloji sözlüğü.....	18
2.4. Tarama ve Sınıflandırma.....	24
2.5. Maximum Entropy.....	26
3. Veri.....	28
3.1. Analiz Verisi.....	28
3.2. Analiz.....	35
3.3. Portföy.....	44
3.4. Bist100 Analiz.....	59
3.5. Kelime Analiz.....	60
4. Sonuç.....	63
Kaynakça.....	65



## Şekil Listesi

Şekil 1 - Fiyat hareketlerinin oluşma aşamaları.....	1
Şekil 2 - Haberlerin toplanması ve analizi.....	8
Şekil 3 - Haber arşiv indirme uygulaması (Örümcek).....	9
Şekil 4 - Haber yorumlayıcı kod mimarisi.....	12
Şekil 5 - Sorgu modeli (Queryable).....	18
Şekil 6 - Türkiye için Ekonomik Psikoloji sözlüğü.....	19
Şekil 7 - Editör ekranı.....	25
Şekil 8 - Editör onaylı haber başarı oranları.....	25
Şekil 9 - Aylık haber dağılımları.....	29
Şekil 10 - Saatlik haber dağılımları.....	30
Şekil 11 - Saatlik haber dağılımları (haberlerin eğilimine göre).....	31
Şekil 12 - Şirket haberleri için karşılaştırmalı aylık dağılımlar.....	33
Şekil 13 - Şirket haberlerinin eğilimlerine göre aylık karşılaştırmalı dağılımı.....	34
Şekil 14 - Genel Haberler Bist100 Getiri (Günlük) Karşılaştırması.....	39
Şekil 15 - Genel Haberler Bist100 Getiri (Saatlik) Karşılaştırması.....	40
Şekil 16 - 4 Gün Gecikmeli Genel Haberler - Bist100 Getiri (Günlük).....	41
Şekil 17 - 4 Saat Gecikmeli Genel Haberler - Bist100 Getiri (Saatlik).....	42
Şekil 18 - Portföy Stratejisi Uygulaması.....	45
Şekil 19 - Bist100 Portföy Stratejisi Getiriler.....	49
Şekil 20 - Garan ve Asyab Portföy Stratejisi Getiriler.....	50
Şekil 21 - Sahol ve Eregl Portföy Stratejisi Getiriler.....	51
Şekil 22 - Portföy Stratejisi I ve II - Akış Şeması.....	55
Şekil 23 - Portföy I - Portföy II Saatlik Eğitim Getiri Karşılaştırma.....	57
Şekil 24 - Bist100 Günlük Fiyat Değişim.....	58
Şekil 25 - Eğitim Tarih Aralığı Ratio ve Kelime Sayıları.....	61
Şekil 26 - Test Tarih Aralığı Ratio ve Kelime Sayıları.....	62

## Tablo Listesi

Tablo 1 - Kelime tanımları.....	19
Tablo 2 - Türkiye Ekonomik Psikoloji Sözlüğü - Negatif Tanımlamalar 1.....	20
Tablo 3 - Türkiye Ekonomik Psikoloji Sözlüğü - Negatif Tanımlamalar 2.....	21
Tablo 4 - Türkiye Ekonomik Psikoloji Sözlüğü - Pozitif Tanımlamalar 1.....	22
Tablo 5 - Türkiye Ekonomik Psikoloji Sözlüğü - Pozitif Tanımlamalar 2.....	23
Tablo 6 - Ham veri, işlenmiş veri ve sözlük bilgileri.....	28
Tablo 7 - Genel haber saatlik dağılımlar (eğilime göre).....	32
Tablo 8 - Asyab Saatlik Haber Dağılımı (eğilime göre).....	33
Tablo 9 - Genel haber kategorisi için Günlük periyot çalışma verisi.....	36
Tablo 10 - Genel haber kategorisi için Saatlik periyot çalışma verisi.....	36
Tablo 11 - Korelasyon Günlük Karşılaştırma.....	37
Tablo 12 - Korelasyon Saatlik Karşılaştırma.....	37
Tablo 13 - Korelasyon Detay (Gecikmeli Haber ve Bist100 Getiri).....	38
Tablo 14 - Saatlik Portföy Stratejileri (Genel Haberler/Bist100).....	47
Tablo 15 - Günlük Portföy Stratejileri (Genel Haberler/Bist100).....	48
Tablo 16 - Portföy-I Test Tarih Aralığı Günlük (Tekil) Getiriler.....	53
Tablo 17 - Portföy II Test Tarih Aralığı Saatlik Getiriler.....	56
Tablo 18 - Bist100 Genel Veriler.....	59
Tablo 19 - Ham ve İşlenmiş Haberlerin İçeriğindeki Kelime Sayıları.....	60
Tablo 20 - Haber Kelime İstatistikleri.....	60

## **Kısaltmalar**

**KH:** Kadir Has

**BOW:** Bag of words

**BIST100:** Borsa İstanbul 100 Endeksi

**TÜFE:** Tüketici Fiyat Endeksi

**C#:** CSharp Yazılım Geliştirme Dili

**LINQ:** Language Integrated Query

**DB:** Database

**LIB:** Library

**Apps:** Applications

**DAL:** Data Access Layer

**BLL:** Business Logic Layer

**UI:** User Interface

**MSHTML:** Microsoft HTML Viewer

**DLL:** Dynamic Link Library

**MSDN:** Microsoft Developer Network

**SQL:** Structured Query Language

**DFSA:** Deterministic Finite State Automata

**URL:** Uniform Resource Locator

**NLP:** Natural Language Processing

**Avg.:** Average

**Correl:** Correlation

## Giriş

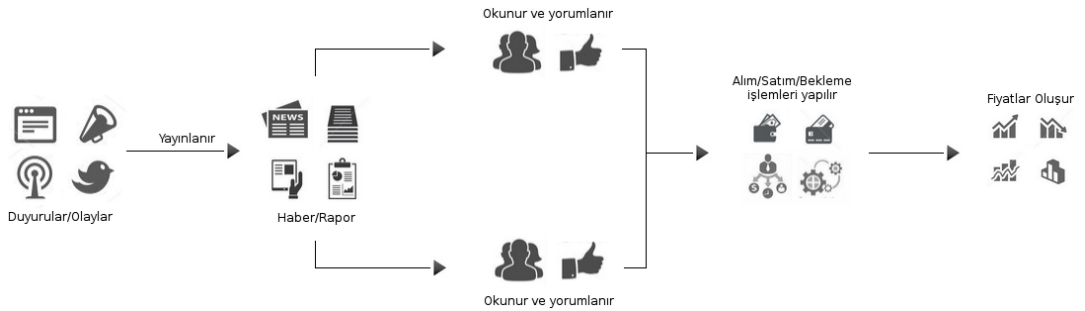
Birikimi olan yatırımcıları ve kaynak ihtiyacı olan girişimcileri bir araya getiren kurumlar, yatırım araçları ve bunların hepsini düzenleyen bir takım kuralların bulunduğu yapıya Finansal Piyasa denir. (Budak & Fıkırkoca, 2011)

Finansal piyasalarda emtia, döviz, hisse senedi, tahviller, pariteler gibi çok sayıda yatırım aracı bulunmaktadır. Yapılan yatırımlar, ileride ne olacağı belirsizliğiyle varlıkların kaybedilme veya değerini yitirme risklerini de beraberinde taşır. Bu risklerin oluşmasında en büyük etkenlerden birisi de fiyat hareketleridir.

Hisse senetlerindeki fiyat değişimlerinde şirketin yatırım, üretim, kâr veya zarar durumları ve politik, ekonomik, ülke bazında veya dünya çapındaki olaylar ile ilgili haberler etkili olabilir. Altın gibi bir emtia grubunda ise politik gelişmeler, ülkeler arası ilişkiler, savaş vb. durumlar fiyat hareketlerini oluşturabilir.

Yatırım araçlarında fiyat hareketlerinin oluşması için Şekil 1’de görüldüğü gibi aşağıdaki adımlar gerekmektedir:

- Bir olayın gerçekleşmesi veya kararların alınması,
- Bu kararların veya gerçekleşen olayların haber kanallarından duyurulması,
- Yatırımcının bu duyuruları okuması ve bilgi düzeyi dâhilinde bir karar vermesi,
- Yatırımcının aldığı karar neticesinde alım veya satım şeklinde tepki vermesi,
- Piyasada fiyatların oluşması (Li, Xie, Chen, Wang, & Deng, 2014).



Şekil 1 - Fiyat hareketlerinin oluşma aşamaları

## Problem

Yatırımcılar ve spekülörler piyasaları analiz ederek daha çok kazanç ve birikim elde etmek amacındadırlar. Bu noktada analiz yapmak için kullanılacak ana kaynak olarak haberler ve makaleler ön plana çıkmaktadır. Bu çalışmada ele alacağımız esas soru ve bunun alt soruları şunlardır: Haberler fiyat hareketlerini ne kadar etkilemektedir?

- Haberler, insan faktörü dışarıda bırakılarak (bilgisayar yardımıyla) pozitif/negatif/nötr olacak şekilde ne kadar başarılı ayrıştırılabilir?
- Ekonomi haberlerinin BIST100 üzerinde, şirket haberlerinin ise kendi fiyat değişimleri üzerinde (saatlik periyot için) ne kadar etkisi vardır?
- Haberler ve fiyatlar arasında bir ilişki var mıdır?
- Buradan yola çıkılarak tahmin yapılabilir mi?
- Yatırım performansını artırmak için başka tür çalışmalar yapılabilir?

## Önceki Çalışmalar

Tarihsel veriler üzerinde sayısal değerler ile elde edilen tahminler geçmişte sınırlı bir başarı sağlamıştır, fakat tahminler yapılırken gerekçeleri gösterilememiştir. Bu noktada, yayınlanan finansal haber ve makalelerin, tahminleri gerekçeleriyle sunacağı yönünde fikir öne sürülmüş ve bu konuda çok sayıda çalışma yapılmıştır (Blasco, Corredor, Rio, & Santamaria, 2005), (Schumaker & Chen, 2009), (Groß-Klußmann & Hautsch, 2011), (Nguyen & Claus, 2013), (Joulin, Lefevre, Grunberg, & Bouchaud, 2008), (Li, Wang, Li, Liu, & Gong, 2014), (Carvalho, Klagge, & Moench, 2011), (Yu, Wu, Chang, & Chu, 2013) (Schumaker, Zhang, Huang, & Chen, 2012), (Fortuny, Smedt, Martens, & Daelemans, 2014), (Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2014).

Başlarda Bag-of-Words (BOW) modeli ile kelime istatistiğine dayalı çalışmalar yapılmış ve başarılı olduğu düşünülmüştür (Fung, Yu, & Lam, 2002). Daha sonraları ise duygu analizi ve öğrenen sistemlerin geliştirilmesiyle birlikte metin içeriklerinin araştırılması psikolojik tanımlamalar düzeyine çekilmiştir. BOW modelinin tek başına yetersiz kaldığı görülmüş olmasına rağmen geliştirilen birçok teknikte BOW ve duygu analizi birlikte kullanılmıştır (Li, Xie, Chen, Wang, & Deng, 2014). Li,

Xie, Chen, Wang'a göre duygu analizi, tahminlerin doğruluğunu artırmakta, fakat tek başına yararlı olmamaktadır.

Bir başka çalışmada BOW kullanılmamış ve sözlük bağımlılığını azaltmak için yeni bir karar destek sistemi (DSS) tasarlanmış, dil işleme ve sınıflandırma yöntemlerinden yararlanılmıştır (Chan & Franklin, 2011).

## **Varsayımlar**

Çalışmanın tutarlılığını sağlamak için haberlerin aynı kaynaktan çıkması gerekmektedir çünkü haberlerin skorlamasında kullanılan kelimeler tanımlanırken yazım dili önem taşımaktadır.

Haber kaynağı internet olduğunda yayınlanma saatinde gecikme olabilecektir (1 saate kadar gecikmeli olabilir).

Fiyatların oluşumunda haberlerin etkisi incelenmektedir. Her ne kadar altın veya döviz de haberlerden etkilenmekteyse de bu çalışmada haberlerin sadece BIST100 ve seçilmiş beş şirketin hisse senedi fiyatı değişimleriyle ilişkisi ele alınacaktır.

Bu çalışmada sadece tahmin becerisi değil, haberlerin doğru bir şekilde pozitif veya negatif olarak puanlanabilmesi de başarı ölçütü olacaktır. Çünkü yatırımcılara yoğun haber akışı içinde bu kapsamda sağlıklı bir değerlendirme sunmak da amaçlanmıştır.

## **1. Haberler**

Haberler yapısı itibari ile pozitif veya negatif bir algı yaratır ya da ilgi odağı dışında kalabilir. Bu konudaki ilk çalışmayı 1965'te Kasper Steiler yapmıştır. Steiler, olayların haber olarak nitelendirilebilmesi için bir takım özellikler taşıması gerektiğini vurgulamıştır; bunlardan bir tanesi "Olumsuzluk" diğer özellikleri ise "Dramatik Öykü" ve "Eve Yakınlık"tır (Rigel, 2000).

Yine 1965'te Johan Galtung ve Mari Ruge bu konuda önemli çalışmalara imza atmışlardır. Buradaki ana unsurlar ise; günlük yaşam, sosyo-kültürel etkiler ve kurumsal yapı'dır. Medya, yapısı gereği çoğunlukla güncel, açık, önemli ve kolay anlaşılabilir olacak olayları ele alır (McQuail, 1987).

## 1.1. Haber Tipleri

Yayınlanan haberler başlıklarına göre şu şekilde gruplandırılabilirler:

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

Genel olarak bu kadar çeşitli başlık içerisinde piyasadaki fiyat hareketlerini etkileyebilecek kategoriler ekonomi, politika ve dünya haberleridir. Ancak, spor grubundan bir hisse senedi için spor haberlerinin de etkisi olabilir. Örneğin Türkiye’de son zamanlarda yaşanan “Şike” davalarında sportif hisseler bu haber grubundan etkilenebilmekteydi. Aynı zamanda ekonomik, politik ve dünya başlıklı haberler tüm yatırım araçlarını kapsayacak genel bir etkiye sahiptir.

### **Politika Haberleri:**

Politik haberler içerisinde, yerel ve dünya ülkelerinde; siyasi aktörlerin söylemleri, kararları ve davranışları yer alır. 2003’te yapılmış olan bir çalışmada, Amihud ve Wohl, Saddam Hüseyin’in devrilmesi ve piyasadaki beklentiler arasındaki ilişkiyi araştırmışlardır. Bu araştırmanın sonuçlarına göre; Saddam Hüseyin’in devrilme olasılığının güçlenmesiyle birlikte pozitif bir hava oluşmuş, hisse senetlerinin yükseldiği, Dolar’ın Euro karşısında güçlendiği ve petrol fiyatlarının düştüğü görülmüştür (Amihud & Wohl, 2003).

### **Dünya Haberleri:**

Dış kaynaklı gelişmelerin düşünülebileceği bu grupta, diğer ülkelerin ekonomik ve başkaca faaliyetleri (uluslararası ilişkiler, askeri, enerji, yatırım, teknoloji vb.) yer alır. Uluslararası şirket birleşme haberlerinin, borsalarda, olumlu etkileri olacaktır. Yine ABD’nin yakın zamanda Parasal Genişleme Programı’nı durdurması, (ilk başlarda tahvil alımlarını azaltması) altın ve gelişmekte olan piyasalar üzerinde negatif bir etki yaratmış, Dolar’ın yerel para birimleri karşısında değer kazandığı görülmüştür.

**Güncel Haberler:**

Genellikle ekonomik değeri olmayan, yaşamdan kesitler içeren ve çoğunlukla Steiler'in de belirttiği gibi olumsuzluk nitelikleri barındıran haber grubudur. Gösteriler, eylemler, kazalar, istatistiki bilgiler, eğitim vb. haberler bu grupta yer alır.

**Ekonomi Haberleri:**

Küresel piyasalar ve buradaki beklentiler, yatırımlar, ticari faaliyetler, ülkelerin cari ve finansal durumları, ithalat, ihracat gibi finansal içerikli haberlerin yer aldığı gruptur.

**Diğer Haberler (Magazin, Spor, Yaşam, Kültür-Sanat):**

Ekonomik bir değeri olmayan, genellikle yaşamdan kesitler içeren haber grubudur. Bu grupta yer alan spor haberleri istisnai bir durum içermektedir. Genel olarak şampiyonluk beklentileri, uluslararası başarılar ve transferler pozitif etki oluştururken, başarısızlıklar negatif yönde etki yapmaktadır.

**1.2. Haberlerin Yatırım Araçları Üzerindeki Etkisi**

Haberlerin yatırım araçları üzerindeki etkisi ile ilgili birçok farklı çalışma mevcuttur. Pozitif ve negatif haberlerin finansal varlıklar üzerindeki etkisi ile ilgili 1999 yılında Veronesi, yaptığı çalışmada, piyasaların iyi zamanlarında kötü haberlerin büyük bir etki yarattığı, ancak, kötü zamanlarda iyi haberlerin zayıf bir etki yarattığı sonucuna ulaşmıştır (Veronesi, 1999).

2005 yılında Boyd, Hu ve Jagannathan, yaptıkları çalışmada, ekonomik genişleme zamanlarında yayınlanan işsizlik verisi haberlerinin hisse senedi fiyatları üzerinde olumlu etki yaratmasına rağmen, ekonomik daralma yaşanırken olumsuz bir etki yarattığını ortaya koymuşlardır. Sermaye risk primi, gelecekteki faiz oranı ile şirket kazançları ve temettü bilgileri farklı ekonomik durumlarda farklı etkiler yaratmaktadır (Boyd, Hu, & Jagannathan, 2005).



### 1.3. Haberlerin İşlenmesi ve Değerlendirilmesi

Daha önce de belirtildiği gibi yatırımcılar haberlerin değerli içeriklerini işleyip, buradan edindikleri fikir ile birtakım kararlar alarak harekete geçmek, daha çok kazanç elde etmek ve zararlardan korunmak istemektedirler.

Burada önemli olan husus, bu işin nasıl yapılacağı noktasındadır. Haber kaynaklarını sürekli takip etmek, yayınlanmış bir haberi gözle okumak ve bu haberin pozitif veya negatif bir anlam taşıdığına karar vermek, hangi yatırım aracı hangi olaylardan etkilenir sorusunu sürekli akılda tutmak... gibi birçok zorluk söz konusudur. Örneğin; “TÜFE’nin beklenen değerden yüksek çıkması borsayı nasıl etkiler?” sorusunun cevabını bilmeyen bir yatırımcı kararsız kalabilir. Haberleri değerlendirirken ne kadar hızlı karar verilirse o kadar çok kazanç veya korunma sağlanacağı da unutulmamalıdır. Aynı anda onlarca farklı haber yayınlandığında ise bunun ne kadar güç ve gecikmeli olabileceği anlaşılmaktadır.

Yapılan çalışmada yazılı uzun bir metnin gözle okunup, pozitif mi negatif mi olduğuna karar verilmesi için geçen sürenin ortalama 30 sn. olduğu görülmüştür (önemli kelimeler işaretlendiği halde). Bilgisayar yardımıyla aynı işlem saniyenin onda biri seviyelerine inebilmektedir. Aynı zamanda haberin başlığı ve verdiği mesaj da zıtlık veya tezat içerebilmektedir. Tecrübesiz bir yatırımcı için ise bu işlemleri gerçekleştirebilmek daha zor olacak ve uzun sürecektir.

Bu çalışmada uygulanan haber toplama ve analiz süreci Şekil 2’de gösterilmiştir.

**Haber arşivi:** Haberler iki farklı ortamdan sağlanmıştır: genel amaçlı bir internet haber portalı ve özel amaçlı bir finansal haber portalı. Bu haberler internet haber portalından otomatik olarak, geliştirilmiş olan bir örümcek yazılım ile, finans portalından da şirket bazındaki haber arşivinin indirilmesi ile elde edilmişlerdir. Bu ham haberler metin dosyaları olarak kaydedilmektedir.

**Ekonomik psikoloji sözlüğü:** Haberlerin ekonomik anlamda değerlendirilmesinde kullanılacak anahtar sözcükler ve nitelikleri (eğilim ve koşul), çalışma kapsamında geliştirilen Türkiye Ekonomik Psikoloji Sözlüğü içerisinde tanımlanmıştır. Şirket bazında haberleri değerlendirmek için de bu sözlüğün yardımıyla bir anahtar sözcük listesi oluşturulmuştur.

**Tarama ve sınıflandırma:** Tüm haberler için, geliştirilen yazılım aracılığıyla anahtar sözcükler taranarak, pozitif/negatif/nötr değerlendirmesi yapılmış ve ilgisiz olanları ayıklanmıştır.

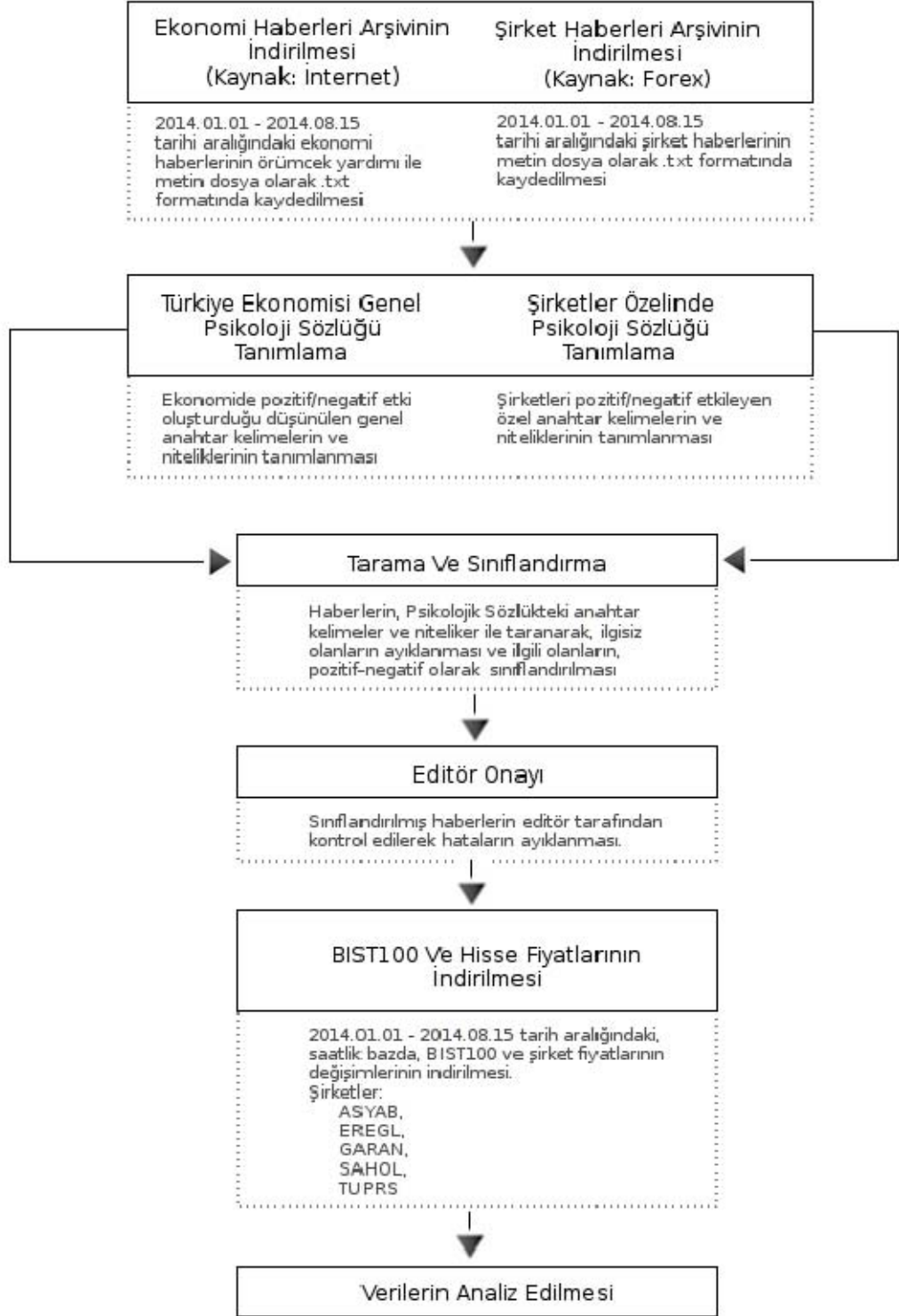
**Editör onayı:** Otomatik olarak yapılan bu değerlendirmelerin performansının ölçümü için haberler tek tek gözle de incelenmiştir. Bu değerlendirmelerin ayrıntıları ileriki bölümlerde verilecektir.

**Fiyat değişimi bilgilerinin indirilmesi:** BIST100 ve 5 adet hisse senedi için saatlik ve günlük periyotlarda, açılış-kapanış fiyatları finansal haber portalı aracılığı ile elde edilmiştir.

Şirket haberleri kategorisinde değerlendirilmek üzere, tahtanın sığ ve derin olma durumu (hacim), piyasa yapıcısı, siyasi ve ekonomik gelişmelerden etkilenme veya etkilenmeme gibi kriterlerin hepsini karşılayabilecek beş adet hisse senedi seçilmiştir.

- Asyab,
- Eregl,
- Garan,
- Sahol,
- Tuprs.

**Verilerin analiz edilmesi:** Yukarıda kısaca değinilmiş olan süreçler sonrasında, işlenmiş veriler üzerinde, haberler ve fiyatlar arasında ilişki olup olmadığı araştırılmıştır.

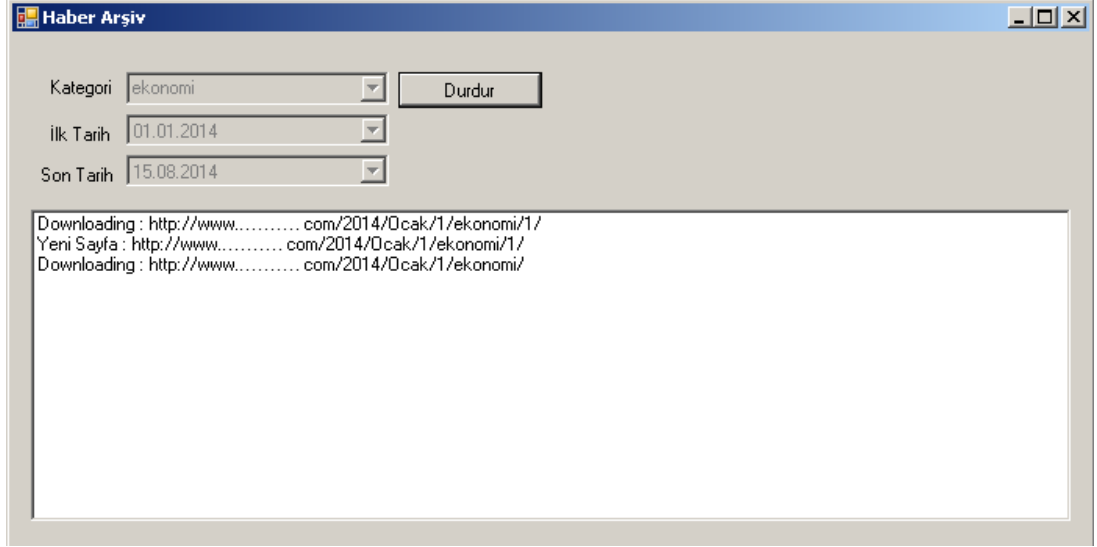


Şekil 2 - Haberlerin toplanması ve analizi

## 2. Yöntem

Haberlerin toplanması ve bilgisayar tarafından yorumlanarak puanlanması için iki ayrı özel uygulama geliştirilmiştir.

- Örümcek (Şekil 3)
- Haber yorumlayıcı (Şekil 4)



Şekil 3 - Haber arşiv indirme uygulaması (Örümcek)

### Örümcek:

Haberler.com internet sitesinden 2014-01-01 ile 2014-12-31 tarihleri arasındaki Türkiye Ekonomi Haberleri arşivinin, metin dosya olarak kaydedilmesi amacıyla yazılmış bir uygulamadır. Bu uygulama yazılırken aşağıdaki kütüphaneler ve yazılım araçlarından faydalanılmıştır:

- HtmlAgilityPack (Obiwan, 2012)
- Yazılım aracı:  
Visual Studio 2010 (C#)
- Gereksinimler:  
Windows işletim sistemi (minimum WindowsXP)  
.NetFramework 4.0 (Microsoft, 2014)

Belirlenen kategoriye göre iki tarih aralığındaki haberler “yyyyMMddHHmm.txt” formatında metin dosya olarak kaydedilir.

İndirilen haberler dört ana kısımdan oluşmaktadır:

**Başlık**, haber içerisindeki olay veya olgular ile ilgili okuyucuya ilk bakışta verilmek istenen mesajın kısa tanımıdır.

**Spot başlık**, sıralamada genellikle başlıktan sonra yer alan, haber ile ilgili ayrıntıların yer aldığı, özet kısımdır.

**Haber metni**, haber ile ilgili detayların yer aldığı ana kısımdır.

**Tarih**, haberin yayınlandığı gün ve saati belirtir.

Uygulama verilen kategori ve tarihe göre tarama işlemine başlayacağı ilk URL'i aşağıdaki şekilde oluşturur.

*http://www.siteadı.com/{0}/{1}/{2}/{3}/*

HtmlAgilityPack kütüphanesi kullanılarak ilk olarak oluşturulan giriş sayfası hafızaya yüklenir ve XPath<sup>1</sup> dili ile sayfada aranmak istenen bölümler tespit edilir.

*XPathExpression = "//\* [contains(@class,'haber\_metni')]"*

Sırasıyla bulunan her bir alt sayfa (en fazla bir derinlik devam edecek şekilde) indirilip haber metinleri kaydedilir. Varsa devam sayfaları da tespit edilerek tüm tarih aralığındaki her güne ait haberler sayfa sayfa indirilir.

Bu çalışmada kullanılmak üzere seçilen 12 aylık periyotta toplam 51.878 adet haber metin dosya olarak kaydedilmiştir. Sitenin indirme hızına bağlı olarak ve aralıklarla çalıştırılarak 8 gün içerisinde tüm haberler kaydedilmiştir.

Aşağıda ana kod parçacığı gösterilmiştir. Alt metodların gövde kısmı gösterilmemiştir.

---

<sup>1</sup> **XPath**: Bir XML dokümanındaki bilgiyi bulmak için kullanılan dildir. Doküman içindeki elemanları ve onlara ait özellikleri incelemeye yarar.

```

class News
{
    public string Category { get; set; }
    public DateTime Date { get; set; }
    public string Title { get; set; }
    public string SpotTitle { get; set; }
    public string Text { get; set; }
}

List<News> newsList = new List<News>();
List<string> mainLinks = new List<string>();

System.ComponentModel.BackgroundWorker bgWorker;

void bgWorker_DoWork(object sender, DoWorkEventArgs e)
{
    DateTime firstDate = dpFirstDate.Value.Date;
    DateTime lastDate = dpLastDate.Value.Date;
    string category = cboCategory.Text;

    mainLinks.Clear();
    newsList.Clear();

    DateTime currDate = firstDate;

    while (currDate <= lastDate)
    {
        if (bgWorker.CancellationPending)
            return;

        string outPageContent = string.Empty;

        string mainPage = GetMainPageLink(currDate, category);
        GetContent(mainPage, out outPageContent);

        int pageCount = GetPageCount(outPageContent);

        for (int page = 1; page <= pageCount; page++)
        {
            if (bgWorker.CancellationPending)
                return;

            string pageContent;
            string pageLink = string.Format("{0}{1}/", mainPage, page);

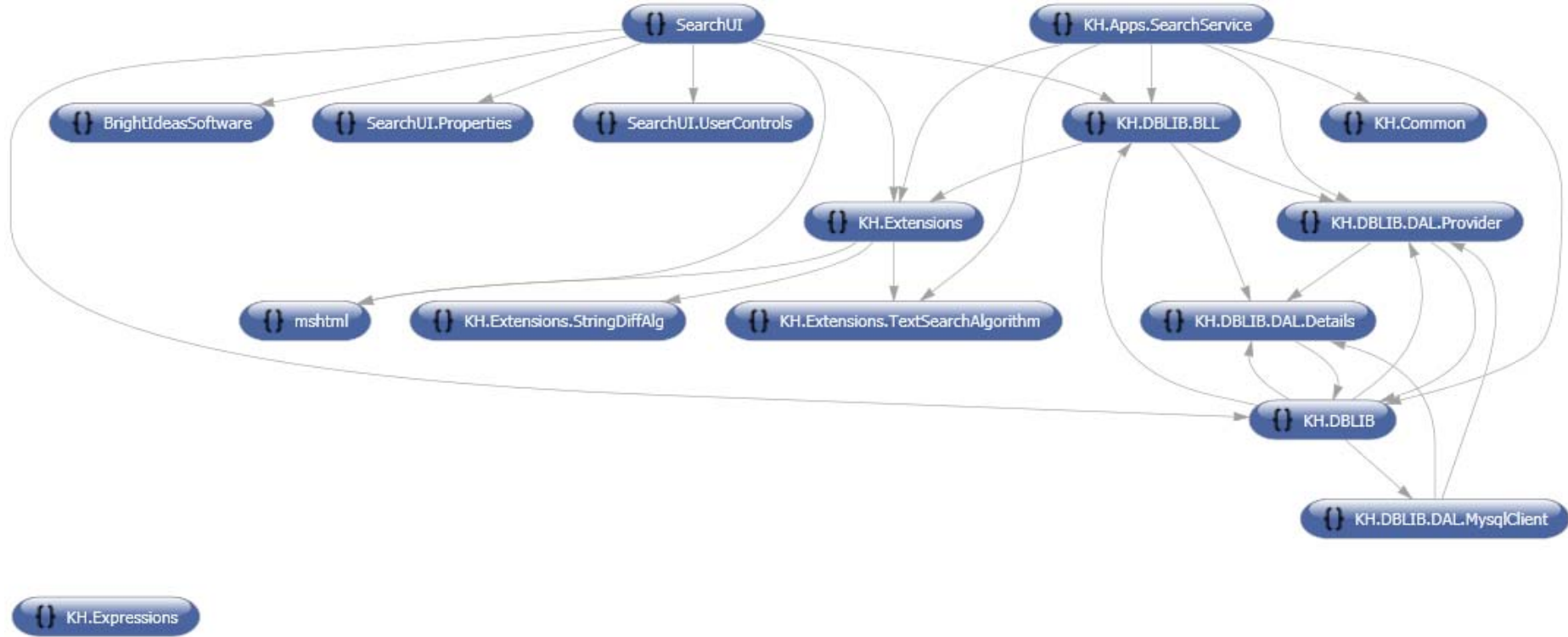
            bool isDown = GetContent(pageLink, out pageContent);

            if (isDown && !string.IsNullOrEmpty(pageContent))
            {
                List<News> pageNews=GetNewsLinks(pageContent,currDate,category)
                newsList.AddRange(pageNews);
            }

            currDate = currDate.AddDays(1);
        }
    }

    private string GetMainPageLink(DateTime date, string cat) {...}
    private bool GetContent(string urlAddress, out string content) {...}
    private int GetPageCount(string content) {...}
    private List<News> GetNewsLinks(string content, DateTime date, string
category) {...}

```



Şekil 4 - Haber yorumlayıcı kod mimarisi

### **Haber yorumlayıcı:**

Metin dosya olarak indirilen haber arşivlerinin teker teker bilgisayar tarafından okunarak, negatif, pozitif veya nötr olarak puanlanması için yazılmış uygulamadır. Bu uygulama yazılırken aşağıdaki algoritma, metod, kütüphaneler ve yazılım araçlarından faydalanılmıştır:

- Yazılım aracı: Visual Studio 2010 (C#)
- Gereksinimler:
  - Windows işletim sistemi (minimum WindowsXP)
  - .NetFramework 4.0 (Microsoft, 2014)
- Aho-Corasick String Search Algorithm C# code (Petricek, 2005)
- Bag-of-Words model (Wikipedia, 2014)
- LINQ (MSDN, 2014)
- BrightIdeasSoftware: ObjectListView (Piper, 2006)
- MSHTML (Microsoft, MSHTML, 2014)
- MySql (Oracle, 2014)

Geliştirilen uygulama mimari olarak üç kısımdan oluşmaktadır (Şekil 4):

- KH.Apps
- KH.DBLIB
- KH.Extensions

KH.Apps projesi, SearchService ve SearchUI katmanlarından oluşmaktadır. **SearchService**, arka planda çalışan, Aho-Corasick algoritması ve LINQ kütüphanesinden yararlanılarak Ekonomik Psikoloji sözlüğünde ön tanımlı kelimelerin otomatik taranıp indekslendiği ve puanlanıp veritabanına kaydedildiği uygulamadır. Sonuçlar MySql veritabanı üzerinde saklanmaktadır. Genel haberler ve tanımlı şirketler için ayrı ayrı tarama yapılmaktadır.

Aşağıda SearchService uygulamasının ana kod parçacığı ve veritabanına kayıt yapıldığı kodlar gösterilmektedir. Alt metodların gövde kısmı gösterilmemiştir.



```

string _startFolder = RemoteConfig.GetValue("StartFolder");
List<Brand> _allBrands =
    Brand.GetBrandList().Where(i => i.IsActive == true).ToList();

void Start(){
    var dirList = _startFolder.GetDirListContainingOnlyFiles("new");
    foreach (var item in dirList)
    {
        Uri u = new Uri(item);

        string segment =
            (u.Segments[u.Segments.Length - 2])
                .ClearCharacters(new char[] { '/' });

        Brand brnd =
            _allBrands.Where(b => b.Name == segment).FirstOrDefault();

        List<BrandKey> kws =
            brnd.GetBrandKeys().Where(k => k.IsActive == true).ToList();

        DirectoryInfo di = new DirectoryInfo(u.AbsolutePath);
        var files = di.GetFiles();

        foreach (var f in files)
        {
            Uri uri = new Uri(f.FullName);
            Document doc = DocumentFactory.New(uri, "text");
            doc.GetText();
            doc.Parse();

            if ((doc == null) || (doc.All == string.Empty)) continue;

            List<BrandKey> allowedKws = doc.FindKeywords(kws);

            if (allowedKws.Count > 0)
                InsertRecord(allowedKws, doc);
        }
    }
}

List<Sign> _allSigns = Sign.GetSignList()
    .Where(i => i.IsActive == true).ToList();

int _signPositive =
    _allSigns.Where(s => s.Name.Contains("+")).FirstOrDefault().ID;

int _signNegative =
    _allSigns.Where(s => s.Name.Contains("-")).FirstOrDefault().ID;

void InsertRecord(List<BrandKey> kws, Document doc)
{
    string tmpAllowed =
        String.Join(",", kws.Select(p => p.ID.ToString()).ToArray());

    List<BrandKey> selfKws =
        kws.Where(p => p.IsSelfSufficient == true).ToList();

    List<Brand> brands =
        _allBrands.AsQueryable().Where(BPredicate(selfKws)).ToList();

    if (selfKws.Count == 0) return;

    var relatedKeys =

```

```

        from d in allowedKws
        from c in brands
        where c.ID == d.BrandID
        group d by d.BrandID into grp
        select new{
            brandID = grp.Key,
            keys = String.Join(",",
                grp.Select(r =>r.ID.ToString()).ToArray()),
            invTypeID = brands.Single(p =>
                p.ID == grp.Key).InvestmentTypeID,
            countOfKeys = grp.Count()
        };

bool isChecked = false;
string autoChkKws = "";

if (relatedKeys.Count() > 0)
{
    isChecked = (kws.Count() ==
        (relatedKeys.Select(p => p.countOfKeys)).Sum());

    autoChkKws = String.Join(",",
        relatedKeys.Select(p => p.keys).ToArray());
}

int countPositive=kws.Where(c => c.SignID == _signPositive).Count();
int countNegative=kws.Where(c => c.SignID == _signNegative).Count();
int ratio = CalculateRatio(countPositive, countNegative);
int opinionId = GetOpinionID(ratio);

int resultID =
    InsertResult(
        doc,
        tmpAllowed,
        (isChecked==false) ? ResultStatus.NONE : ResultStatus.AUTO,
        isChecked,
        autoChkKws,
        ratio
    );

foreach (var item in relatedKeys)
{
    ResultConfirm.InsertResultConfirm(
        resultID,
        item.invTypeID,
        item.brandID,
        "," + item.keys + ",",
        opinionId);
}

ResultText.InsertResultText(resultID, doc.PureText);
}

static Expression<Func<Brand, bool>>BPredicate(List<BrandKey> keys){...}
int GetOpinionID(int ratio) {...}
int CalculateRatio(decimal countPositive, decimal countNegative) {...}

```

**SearchUI**, tanımlamaların ve rapor ekranlarının bulunduğu görsel uygulamadır. Yatırım araçları, fikir/düşünce, marka, kelime tanımlamaları ve rapor (editör) ekranlarını içermektedir. Editör penceresinde (Şekil 7), bulunan kelimeler mavi renkle boyanarak (boyama işlemlerinde mshtml<sup>2</sup> kütüphanesinden faydalanılmıştır) kullanıcının kelimeleri kolayca görebilmesi amaçlanmıştır. Yine bu ekranda haberlerin liste olarak ekranda gösterilmesi için BrightIdeasSoftware tarafından geliştirilen ObjectListView'dan yararlanılmıştır (Piper, 2006).

KH.DBLIB, yukarıda bahsedilen uygulamaların kullandığı nesnelere sağlayan ve veritabanı ile iletişim kurduğu katmandır. Bu kütüphane iki alt katmandan oluşur: DAL (Data Access Layer), veritabanı ile nesnelere doğrudan erişim sağladığı katmandır. BLL (Business Logic Layer), uygulamalar ve DAL arasındaki katmandır. Uygulamaların veritabanına doğrudan erişimini engellemek için kullanılır.

KH.Extensions, algoritmaların ve ortak kod parçacıklarının yer aldığı katmandır.

## 2.1. Aho-Corasick Dizi Arama Algoritması

Kullanıcı tarafından belirtilen bir kelime veya kelime dizisinin daha büyük kelimeler dizisi (metin) içerisinde aranması konusu, bilgi alma ve metin düzenleme uygulamalarının temelini teşkil etmektedir. Arama algoritmaları temel olarak iki başlık altında sınıflandırılmaktadır:

- Tek desenli arama algoritmaları (single pattern):  
Boyer-Moore (literatürde karşılaştırmalar için sıklıkla kullanılır),  
Knutth-Morris-Pratt
- Çoklu desenli arama algoritmaları (multiple pattern):  
Aho-Corasick,  
Commentz-Walter,  
Karp-Rabin

Tek desenli algoritmalarda, verilen her bir kelime veya kelime dizisi için tüm metin üzerinde ayrı ayrı tarama yapılır. Çoklu desenli algoritmalarda ise verilen tüm kelime veya kelime dizisi metin üzerinden tek geçişte yakalanır.

---

<sup>2</sup> **mshtml**: Windows Internet Explorer tarayıcısının ana HTML bileşenidir ve başka uygulamalar ile de kullanılabilir.

Yaygın olarak kullanılan yöntem olasılığa dayalı olmayan sonlu durum otomat<sup>3</sup> (DFSA: deterministic finite state automata) metodudur. Bu yöntemde metin içerisindeki her bir karakterin tek tek incelenmesi gerekmektedir (Fan & Su, 1994).

Aho-Corasick, desenlerin yapılandırılması DFSA üzerine kurulmuş çoklu desenli bir arama algoritmasıdır. Alfred V. Aho ve Margaret J. Corasick tarafından 1975 yılında geliştirilmiştir. Üç aşamada diziler aramaya hazır hale getirilir:

- Desen ağacı oluşturma
- Dizinin en uzun son ekinin bulunması
- Sonuç

Desen ağacı oluşturulurken aranmak istenen tüm dizilerin (kelimelerin) her biri sırasıyla taranır. Dizilerin ilk karakterleri genellikle başlangıç setini oluşturur. Her bir dizinin alt karakter kümesi için bulunan sonuçlar başlangıç karakter setine bağlanır. Bu şekilde tüm kelimeler birbirinin alt kümesi olabiliyorsa ilişkilendirilmiş olur. Desen ağacı oluşturulduktan sonra, karakterlerin tekrarının önlenmesi için en uzun son ek kontrolü yapılır. Son adımda ise her bir setin arkasına sonuç eklenir (Hasib, Motwani, & Saxena, 2013).

Aho-Corasick algoritması, Unix komutlarından fgrep'in alt yapısını oluşturmaktadır. Bilgisayar dünyası, biyoloji bilimi (DNA yapısının çözümlenmesi) gibi birçok alanda da kullanılmaktadır.

Büyük bir metin içerisinde çok sayıda dizi aranıyorsa bu algoritma en iyi sonucu vermektedir. Ancak, az sayıda dizi aranıyorsa, desen ağacı oluşturma aşaması zaman aldığı için indexOf<sup>4</sup> yöntemi tercih edilmelidir. Yapılan hız testinde, bir haber metni içerisinde 100000 kelime (AND, OR ve NOT koşulları ile birlikte 500000) indexOf yöntemi ile arandığında yaklaşık 10 sn sürerken, Aho-Corasick ve indexOf yöntemi birlikte kullanıldığında 1.23 sn sürmektedir. Sadece Aho-Corasick ile koşullar dışarıda bırakılarak yapılan testte ise metnin 0.023 sn'de tarandığı görülmüştür.

---

<sup>3</sup> **Otomat(Automata):** Özdevinir, kendini tekrar eden.

<sup>4</sup> **IndexOf:** Belirtilen karakterin dizi içerisinde ilk geçtiği dizini bulan metod ismi

## 2.2. Bag-of-Words Model Tabanlı Genişletilmiş Sorgulama

Bir metnin bilgisayar tarafından yorumlanması için günümüzde kullanılan en yeni teknik duygu analizidir. Duygu analizi yapılırken doğal dil işleme yöntemleri, metin analizi, otomatik öğrenme teknikleri bir arada kullanılabilir.

Bag-of-Words ise doğal dil işlemede kullanılan bir modeldir. Daha çok metin sınıflandırmada kullanılır. Bu modelde dil bilgisi kurallarının ve kelimelerin sırasının bir önemi yoktur. Kelimelerin kaç adet geçtiği ile ilgilenilir. Bu modelden ilk olarak, 1954'te Zellig Harris, "Distributional Structure" isimli makalesinde söz etmiştir (Harris, 1954).

Bu çalışmada Bag-of-Words modeli geliştirilerek dil işleme yöntemleri, otomatik öğrenme gibi teknikleri kullanmadan bir sonuç elde etmek amaçlanmıştır. Bunun için modele daha önce kullanılmamış olan Sorgulama (Queryable) yeteneği eklenmiştir. Bu eklenti ile birlikte AND/OR mantıksal operatörleri kullanılabilir, herhangi bir ifade/kelimenin metin içerisinde geçmemesi koşulu koyulabilir (NOT), yine herhangi bir ifade/kelimenin diğer koşullardan uzaklığına bakılabilir. Aynı zamanda kelime kökleri tanımlanabilir. Böylece dil bilgisi kurallarına tabi olmadan sadece kelime tanımlayarak çalışılabilir (Şekil 5).



Şekil 5 - Sorgu modeli (Queryable)

## 2.3. Türkiye Ekonomik Psikoloji sözlüğü

Metinler üzerinde duygu analizi çalışmaları yapılırken dil bilgisi ve psikoloji sözlüklerinden yararlanılmaktadır. Loughran and McDonald Financial Sentiment Dictionaries (McDonald, 2013) ve Harvard Psychological Dictionary (Hurwitz, 2002) örnek olarak verilebilir. Türkçe doğal dil işleme konusunda Zemberek-NLP<sup>5</sup> gibi kütüphaneler bulunmaktadır, ancak, ekonomik alanda Türkçe için hazırlanmış bir sözlüğe rastlanmamıştır.

<sup>5</sup> **Zemberek-NLP:** Türkçe için geliştirilmiş doğal dil işleme kütüphanesidir. (<http://zembereknlp.blogspot.com.tr/>)

Yukarıda anlatılmış olan sorgu model teknikle birlikte Türkiye Ekonomik Psikoloji Sözlüğü'nün ilk adımı atılmıştır. Kelimelerin doğru tanımlanabilmesi için ekonomi ve piyasa veya şirketler hakkında bilgi sahibi olunması, faaliyet alanlarının bilinmesi gibi bazı zorluklar bulunmaktadır. Bu noktada büyük bir ön araştırma yapılmıştır. Sözlük hazırlanırken yaklaşık 35.000 haber gözden geçirilerek yazım dili hakkında bilgi edinilmiştir. İlk tanımlamaların yapılması yaklaşık iki hafta sürmüştür. İlk tanımlamadan sonraki aşamada iki defa optimizasyon yapılmış olup yaklaşık bir hafta da düzeltmeler için harcanmıştır. Daha çok kelime kökleri ve tamlamalar kullanılırken, birçok tanım için hem pozitif, hem negatif ve bazı gözlemler yapmak için de nötr tanımlamalar eklenmiştir. Genel ekonomi haberleri için ayrı, her bir şirket için de ayrı ayrı kütüphane oluşturulmuştur. Hazırlanmış olan sözlük kullanılarak yaklaşık 6.000 haber içeriği editör tarafından gözle okunarak onaylanmıştır. Sonuç olarak genel ortalama %91'lik bir başarı sağlandığı gözlemlenmiştir.

EffectType	Sign	BrandName	SearchKey	AndKey	NotKey	AutoApprove	Language	Ratio
Tüm Yatırım Araçları	Negatif (-)	ALL	CEMAAT	HÜKÜMET		True	Türkçe	50
Tüm Yatırım Araçları	Pozitif (+)	ALL	TAKASLA TİCARET			True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	ÖTV	ARTIR*YÜKSEL*ARTTI*AR...		True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	POZİSYON AÇIĞI	ARTIR*YÜKSEL*ARTTI*AR...		True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	ZAYIF	EKONOMİ15		True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	DIŞ BÖRÇ	ARTIR*YÜKSEL*ARTTI*AR...		True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	FAALİYET DUR			True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	HİZMET SEKTÖR	DÜŞTÜ*AZAL*DIP*GERİ*ZA...	DIPLOMA*DIPLOM...	True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	ANLAŞMA	SAĞLANAMA10*SAĞLAYA...		True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	DARBE			True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	RISK İŞTAH	AZAL*DÜŞÜŞ*DÜŞTÜ		True	Türkçe	50
Tüm Yatırım Araçları	Pozitif (+)	ALL	RISK İŞTAH	ARTIR*YÜKSEL*ARTTI*AR...		True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	PARA ARZI	DÜŞTÜ*AZAL*DIP*GERİ*ZA...	DIPLOMA*DIPLOM...	True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	SANAYİ CİRD	DÜŞTÜ*AZAL*DIP*GERİ*ZA...	DIPLOMA*DIPLOM...	True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	BELİRSİZLİK	DEVAM*HAKİM*EYLEM		True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	İŞİD			True	İngilizce	50
Tüm Yatırım Araçları	Negatif (-)	ALL	İHRACAT	AZAL*DÜŞÜŞ*DÜŞTÜ*GER...	GERİ SAYIMI30	True	Türkçe	50
Tüm Yatırım Araçları	Negatif (-)	ALL	ÜRETİM	DURDU*ARA VERDİ*ARA V...		True	Türkçe	50
Tüm Yatırım Araçları	Pozitif (+)	ALL	KURTARMA PROGRAMI	ÇIKTI25*ÇIKAL25*BAŞLA		True	Türkçe	50
Tüm Yatırım Araçları	Pozitif (+)	ALL	İSTİHDAM	REKOR*ARTTI*YÜKSELDİ		True	Türkçe	50

## Şekil 6 - Türkiye için Ekonomik Psikoloji sözlüğü

Bir kelime tanımlanırken etki alanı (tüm yatırım araçları / tek yatırım aracı), işaret (pozitif / negatif / nötr), marka (hepsi / şirket), kilit kelime, AND koşulu, NOT koşulu, dili, onay şekli (otomatik / elle), kelime başarı ortalaması (başlangıç değeri: 50) bilgileri girilmektedir (Şekil 6).

Tablo 1 - Kelime tanımları

Kelime Tanımları							
	Genel	Ereğl	Garan	Sahol	Tuprs	Asyab	Toplam
Nötr	12	0	0	0	0	0	12
Negatif	152	59	45	46	57	45	404
Pozitif	106	44	43	38	51	29	311

Tablo 2 - Türkiye Ekonomik Psikoloji Sözlüğü - Negatif Tanımlamalar 1

Etki Alanı	Eğilim	Marka	Anahtar Kelime	AND Koşulu	NOT Koşulu
Tüm Yatırım Araçları	Negatif (-)	ALL	KREDİ NOTU	DÜŞTÜ*AZALT*DÜŞÜR*KALKTI*KALDIR	
Tüm Yatırım Araçları	Negatif (-)	ALL	FAİZ	ARTIR*YÜKSEL*ARTTI*ARTIŞ*ÇIKTI*BASKI*YÜKSEK	
Tüm Yatırım Araçları	Negatif (-)	ALL	SAVAŞ	IRAK*SURİYE*İSRAİL*MISIR*İÇ SVAŞ	
Tüm Yatırım Araçları	Negatif (-)	ALL	İŞİD		
Tüm Yatırım Araçları	Negatif (-)	ALL	17 ARALIK	2013	
Tüm Yatırım Araçları	Negatif (-)	ALL	CARİ AÇI	ARTIŞ*ARTTI*ÜZERİ*YÜKSEK*YÜKSEL	
Tüm Yatırım Araçları	Negatif (-)	ALL	İŞSİZLİK	ARTTI*ARTIŞ*ÇIKTI*YÜKSEK*ARTIYOR*YÜKSEL	
Tüm Yatırım Araçları	Negatif (-)	ALL	TARIM DIŞI İSTİHDAM	AZALDI*AZALİYOR	
Tüm Yatırım Araçları	Negatif (-)	ALL	SİYASİ BELİRSİZLİK	ORTAM*BAŞLA	
Tüm Yatırım Araçları	Negatif (-)	ALL	MOODY'S	NOT*AZALT*DÜŞÜR*NEGATİF*DURAGAN*OLUMSUZ *İNDİRİM*UYAR*DURAGAN	
Tüm Yatırım Araçları	Negatif (-)	ALL	KÜRESEL PİYASA	DÜŞÜŞ*NEGATİF*SATIŞ*AŞAĞI*YAVAŞ*GERİ*OLUMSUZ	
Tüm Yatırım Araçları	Negatif (-)	ALL	ENFLASYON	ARTTI*ARTIŞ*YÜKSEL*ÜZERİ*YÜKSEK	FARK 15*MAAŞ 15
Tüm Yatırım Araçları	Negatif (-)	ALL	ZORUNLU KARŞILIK	ARTTI*ARTIŞ*ARTIR	
Tüm Yatırım Araçları	Negatif (-)	ALL	KÜRESEL	KRİZ*KRİTİK*OLUMSUZ	
Tüm Yatırım Araçları	Negatif (-)	ALL	GÖRÜNÜM	NEGATİF*DURAGAN*DÜŞÜR*DÜŞTÜ*AŞAĞI*BOZUL *KÖTÜ*OLUMSUZ	
Tüm Yatırım Araçları	Negatif (-)	ALL	SANAYİ ÜRETİM	DÜŞTÜ*AZAL*GERİ*ZAYIF*DÜŞÜK*DÜŞÜŞ *NEGATİF	GERİ SAYIM
Tüm Yatırım Araçları	Negatif (-)	ALL	DARALMA	EKONOMİ*PİYASA	
Tüm Yatırım Araçları	Negatif (-)	ALL	PARALEL	DEVLET*YAPI*MEDYA	ŞEKİL 10*OLARAK 10 *BİÇİMDE 10*LİK 8*İNDE 8

Tablo 3 - Türkiye Ekonomik Psikoloji Sözlüğü - Negatif Tanımlamalar 2

Etki Alanı	Eğilim	Marka	Anahtar Kelime	AND Koşulu	NOT Koşulu
Tüm Yatırım Araçları	Negatif (-)	ALL	CEMAAT	HÜKÜMET	
Tüm Yatırım Araçları	Negatif (-)	ALL	RİSK İŞTAH	AZAL*DÜŞÜŞ*DÜŞTÜ	
Tüm Yatırım Araçları	Negatif (-)	ALL	MAZOT	ARTIR*YÜKSEL*ARTTI*ARTIŞ*ÇIKTI*ZAM	
Tüm Yatırım Araçları	Negatif (-)	ALL	VERGİ	ZAM*ARTTI*YÜKSEL*ARTIŞ*YÜKÜ*YÜKSEK*BORÇ *İNCELEME*ZARAR 15*BORC 15	DÜŞÜK
Tüm Yatırım Araçları	Negatif (-)	ALL	DARBOĞAZ	EKONOMİ	
Tüm Yatırım Araçları	Negatif (-)	ALL	RESESYON		
Tüm Yatırım Araçları	Negatif (-)	ALL	VARLIK ALIM	AZALT*SON VER*BİTİRDİ*BİTİRECE	
Tüm Yatırım Araçları	Negatif (-)	ALL	İMALAT ENDEKSİ	DÜŞÜK*ALTINDA*ZAYIF*DÜŞTÜ*GERİ	GERİ SAYIM
Tek Yatırım Aracına Özel	Negatif (-)	ASYAB	SORUŞTURMA		
Tek Yatırım Aracına Özel	Negatif (-)	ASYAB	OPERASYON		OPERASYONEL
Tek Yatırım Aracına Özel	Negatif (-)	EREGL	YABANCI YATIRIM	SATIŞ*ÇIKIŞ*ÇIKIYOR	
Tek Yatırım Aracına Özel	Negatif (-)	EREGL	ARCELORMİTTAL	HİSSE SAT*SATIŞ	
Tek Yatırım Aracına Özel	Negatif (-)	EREGL	ÇED SÜRECİ	BAŞLA	
Tek Yatırım Aracına Özel	Negatif (-)	EREGL	HAMMADDE GİDER	YÜKSEK*PAHALI	
Tek Yatırım Aracına Özel	Negatif (-)	GARAN	HEDEF FİYAT	NEGATİF*AŞAĞI*AZALT*İNDİ*DÜŞÜR*OLUMSUZ	
Tek Yatırım Aracına Özel	Negatif (-)	GARAN	ZORUNLUK KARŞILIK	ARTIR*ARTTI*YÜKSEL*ÇIKAR	
Tek Yatırım Aracına Özel	Negatif (-)	GARAN	BATIK KREDİ		
Tek Yatırım Aracına Özel	Negatif (-)	GARAN	MEVDUAT ÇIKIŞ	ARTIR*ARTTI*YÜKSEL	
Tek Yatırım Aracına Özel	Negatif (-)	TUPRS	RAFİNERİ MARJİ	EKSI*DÜŞÜK	



Tablo 4 - Türkiye Ekonomik Psikoloji Sözlüğü - Pozitif Tanımlamalar 1

Etki Alanı	Eğilim	Marka	Anahtar Kelime	AND Koşulu	NOT Koşulu
Tüm Yatırım Araçları	Pozitif (+)	ALL	MOODY'S	NOT*ARTIR*ARTIŞ*OLUMLU*POZİTİF	
Tüm Yatırım Araçları	Pozitif (+)	ALL	EKONOMİ	İYİLEŞME*OLUMLU*POZİTİF*DÜZEL*İLİMLİ*İYİMSER	ZAYIF 30
Tüm Yatırım Araçları	Pozitif (+)	ALL	ÇÖZÜM SÜRECİ		BAŞARISIZ 30*BİTT 30*SONA ER 30
Tüm Yatırım Araçları	Pozitif (+)	ALL	CARİ AÇI	AZAL*DÜŞÜŞ*DÜŞTÜ*GERİLE*ALTINDA	
Tüm Yatırım Araçları	Pozitif (+)	ALL	İŞSİZLİK	AZALDI*AZALIYOR*İNDİ*DÜŞÜK*GERİ	GERİ SAYIM
Tüm Yatırım Araçları	Pozitif (+)	ALL	TARIM DIŞI İSTİHDAM	ARTIŞ*ARTTI	
Tüm Yatırım Araçları	Pozitif (+)	ALL	VERGİ	BARİŞ*AF*KALDIRIL*İNDİ*HAFİFLE*SİLİN*MUAFİYET	
Tüm Yatırım Araçları	Pozitif (+)	ALL	SİYASİ BELİRSİZLİK	KALKTI*BİTTİ	
Tüm Yatırım Araçları	Pozitif (+)	ALL	İTHALAT	AZALDI*AZALIYOR*DÜŞTÜ*DÜŞÜŞ	
Tüm Yatırım Araçları	Pozitif (+)	ALL	KÜRESEL PİYASA	POZİTİF*ARTIŞ*ARTTI*ALIM*YUKARI*OLUMLU	
Tüm Yatırım Araçları	Pozitif (+)	ALL	ENFLASYON	DÜŞÜŞ*NEGATİF*AZAL*DÜŞTÜ*ALTINDA	FARK 15*MAAŞ 15
Tüm Yatırım Araçları	Pozitif (+)	ALL	ZORUNLU KARŞILIK	DÜŞTÜ*AZALT*DÜŞÜR*KALKTI*KALDIR	
Tüm Yatırım Araçları	Pozitif (+)	ALL	BAYRAM HAVASI	PIYASA*EKONOMİ	
Tüm Yatırım Araçları	Pozitif (+)	ALL	GÖRÜNÜM	POZİTİF*ARTIŞ*ARTTI*YUKARI*OLUMLU	
Tüm Yatırım Araçları	Pozitif (+)	ALL	AKARYAKIT	DÜŞTÜ*DÜŞÜŞ*İNDİRİM	
Tüm Yatırım Araçları	Pozitif (+)	ALL	FAİZ	İNDİ*DÜŞTÜ*DÜŞÜŞ*AZALT*TEK HANE	
Tüm Yatırım Araçları	Pozitif (+)	ALL	GÜVEN ENDEKSİ	YÜKSEK*YÜKSEL*ZİRVE*ARTTI*ÇIKTI	
Tüm Yatırım Araçları	Pozitif (+)	ALL	BEKLENTİ	ARTTI*ARTIŞ*POZİTİF*YÜKSEL	
Tüm Yatırım Araçları	Pozitif (+)	ALL	SIKILAŞTIRMA	PARA*DEVAM	
Tüm Yatırım Araçları	Pozitif (+)	ALL	ENERJİ FİYAT	DÜŞTÜ*DÜŞÜŞ*İNDİ*AZALT*GERİ	GERİ SAYIM

Tablo 5 - Türkiye Ekonomik Psikoloji Sözlüğü - Pozitif Tanımlamalar 2

Etki Alanı	Eğilim	Marka	Anahtar Kelime	AND Koşulu	NOT Koşulu
Tüm Yatırım Araçları	Pozitif (+)	ALL	SANAYİ ÜRETİM	YÜKSEL*ARTTI*ARTI*ÇIKTI*ZİRVE*YÜKSEK*POZİTİF	
Tüm Yatırım Araçları	Pozitif (+)	ALL	DIŞ TALEP	CANLAN*ARTTI*ARTIŞ*YUKARI*YÜKSEL	
Tüm Yatırım Araçları	Pozitif (+)	ALL	İLİMLİ	EKONOMİ*POLİTİKA*PIYASA	
Tüm Yatırım Araçları	Pozitif (+)	ALL	İMALAT ENDEKSİ	YÜKSEK*YÜKSEL*ARTTI*ÇIKTI*GÜÇLÜ	
Tüm Yatırım Araçları	Pozitif (+)	ALL	RİSK İŞTAH	ARTIR*YÜKSEL*ARTTI*ARTIŞ*ÇIKTI	
Tüm Yatırım Araçları	Pozitif (+)	ALL	İHRACAT	YÜKSEL*ARTTI*ARTI*ÇIKTI*POZİTİF*YÜKSEK*REKOR	
Tüm Yatırım Araçları	Pozitif (+)	ALL	BÜTÇE	FAZLA*ARTIŞ	
Tüm Yatırım Araçları	Pozitif (+)	ALL	BARIŞ SÜRECİ	DEVAM 30*OLUMLU*POZİTİF*İLİMLİ*İYİMSER*KATKI	BAŞARISIZ 30*BİTT 30*SONA ER 30
Tek Yatırım Aracına Özel	Pozitif (+)	ASYAB	ÖDENMİŞ SERMAYE	ARTIR*ARTTI*YÜKSEL*ÇIKAR	
Tek Yatırım Aracına Özel	Pozitif (+)	ASYAB	NET DÖNEM KÂR		
Tüm Yatırım Araçları	Pozitif (+)	EREGL	EREGL	ORTAK*SATIŞ*TEMETTÜ*KÂR PAYI*İŞLEME AÇ*GENEL KURUL*BİRLEŞME*NET KÂR*SERMAYE ART	BİRİNCİ SEANS KAPANIŞ*İLK SEANS*İKİNCİ SEANS KAPANIŞ
Tek Yatırım Aracına Özel	Pozitif (+)	EREGL	BEDELSİZ	DAĞIT*KARAR*AÇIKLA*VERECE*PAY	
Tek Yatırım Aracına Özel	Pozitif (+)	EREGL	SOĞUK HADDE	KURMA*ÇALIŞMA*KAPASİTE	
Tek Yatırım Aracına Özel	Pozitif (+)	GARAN	REKOR KÂR		
Tek Yatırım Aracına Özel	Pozitif (+)	GARAN	KOMİSYON	GELİR*YÜKSEK*GÜÇLÜ*ARTAN	
Tek Yatırım Aracına Özel	Pozitif (+)	SAHOL	ENERJİSA	KAPASİTE ART	
Tek Yatırım Aracına Özel	Pozitif (+)	SAHOL	PAZAR PAYI	ARTTI*ARTIŞ*ÇOĞAL*YÜKSEL*ARTI*YUKARI	
Tek Yatırım Aracına Özel	Pozitif (+)	TUPRS	TEŞVİK		
Tek Yatırım Aracına Özel	Pozitif (+)	TUPRS	RESİD UPGRADING		

Hazırlanmış olan sözlüğün ağırlıklı olarak negatif tanımlardan oluştuğu göze çarpmaktadır. Sadece Genel kategorisi için nötr tanımlama girilmiştir (Tablo 1). Bunun nedeni ilgili kelimenin daha çok hangi sınıf haberler (pozitif/negatif) içerisinde yer aldığını gözlemlemektir. Buna örnek olarak Almanya başbakanı Merkel'i ve yine Türkiye'den örnek olarak Ali Babacan'ı gösterebiliriz. Merkel için pozitif ve negatif oranlarının birbirine çok yakın çıktığı görülmüştür. Babacan geçen haberlerin %90 üzerinde pozitif yönlü olduğu tespit edilmiştir. Bu da piyasanın Babacan'ı neden kabinde görmek istediğinin yanıtını oluşturmaktadır.

Nötr tanımlamalar, puanlama yapılırken hesaba katılmamakla birlikte nasıl bir etki göstereceği bilinemeyen kelimeler için veya sadece istatistik amaçlı da kullanılabilir.

#### 2.4. Tarama ve Sınıflandırma

Haber arşivinin örümcek yardımıyla indirilmesi ve sözlüğün oluşturulması aşamalarından sonra, Haber Yorumlayıcı uygulaması yardımıyla otomatik olarak sınıflandırılması ve sınıflandırma sonucunun doğruluk oranının saptanması amacıyla editor kontrolünden geçirilmesi sağlanmıştır.

Otomatik yorumlama/puanlama sırasında nötr kelimeler hesaba katılmadan, sadece negatif ve pozitif kelimelerin ağırlıklı ortalamaları üzerinden haberlerin işareti hesaplanmıştır.

$$pozitifPuan = \left( \frac{pozitifKelimeSayısı}{negatifKelimeSayısı + pozitifKelimeSayısı} \right) * 100$$

$$negatifPuan = 100 - \left( \frac{negatifKelimeSayısı}{negatifKelimeSayısı + pozitifKelimeSayısı} \right) * 100$$

Puanlamalar:

- 0-49: Negatif
- 50: Nötr
- 51-100: Pozitif

Daha sonraki çalışmalarda her bir kelime için başarı oranları tutularak, editör onayına getirmeden otomatik onay mekanizması eklenmesi düşünülmektedir.

Begin Date	Investment Type	Make Group
24.01.2013	Equity	
End Date	Brand	Match Whole Word
24.01.2015	ALL	Match Case
Get Records		

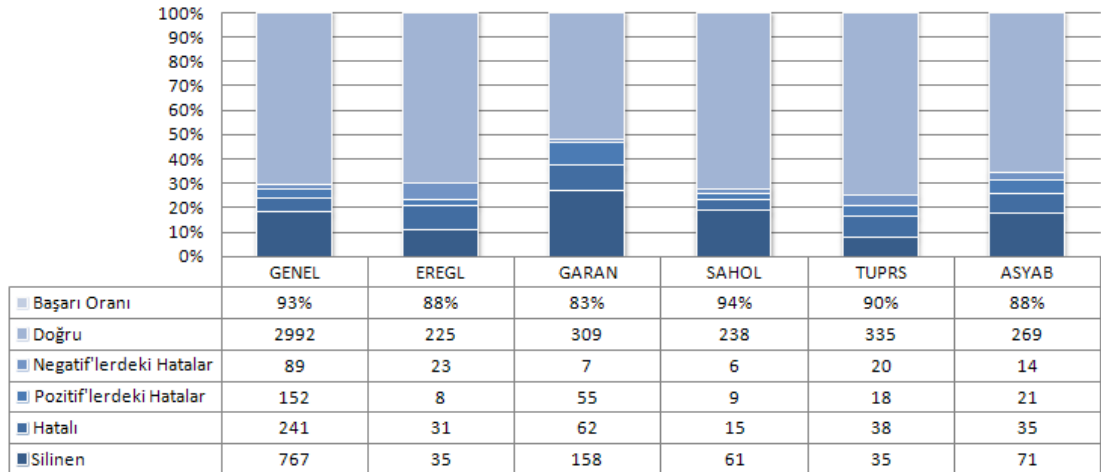
  

Line	TransDate	TransTime	Title	Ratio	Status
2220	2014-01-30	15:26:00	Tüketici Kredileri ve Kredi Kartı...	50	AUTO
2221	2014-01-30	15:30:00	Bankacılık Sektörü Mevduatları Arh...	100	AUTO
2222	2014-01-30	15:34:00	"Sermaye Hareketlerine Herhangi Bir...	59	AUTO
2223	2014-01-30	15:35:00	Shell Akademi Projesini Durdurmayı...	50	AUTO
2224	2014-01-30	15:35:00	Sijansi ve Ekonomik İstikrar Oturusa...	52	AUTO
2226	2014-01-30	15:40:00	Geçmiş Günümüze Kamu Maliyesi" Paneli	58	AUTO
2227	2014-01-30	15:43:00	Bursa'da Otonom Test Merkezi Meray...	50	AUTO
2228	2014-01-30	15:44:00	Para Arzı 960.5 Milyar TL'ye Yükseldi	50	AUTO
2229	2014-01-30	15:51:00	Maliye Bakanı: Sıyirt'a 25 Milyon...	57	AUTO
2230	2014-01-30	15:54:00	Merkez Bankası Rezervleri 1 Haftada 4.5...	0	AUTO
2231	2014-01-30	15:59:00	Geçmiş Günümüze Kamu Maliyesi" Paneli	75	AUTO
2232	2014-01-30	16:06:00	Alınur Aktar: İnegöl Mobilyası...	50	AUTO
2233	2014-01-30	16:14:00	Bankacılık Sektörü Kredi Hacmi 1 Haftada...	100	AUTO
2234	2014-01-30	16:14:00	Tüketici Kredileri 248.9 Milyar TL'ye...	100	AUTO
2235	2014-01-30	16:15:00	Piyasalarda Son Durum	57	AUTO
2236	2014-01-30	16:20:00	Ticmb Döviz Kurları	50	AUTO
2238	2014-01-30	16:25:00	Geçmiş Günümüze Kamu Maliyesi" Paneli	60	AUTO
2239	2014-01-30	16:26:00	TL'ye Zaman Kazan	63	AUTO
2240	2014-01-30	16:29:00	Ekonomik İstikrarsızlığın Faturasını...	56	REVER...
2241	2014-01-30	16:34:00	Bankalararası Dahil Toplam Mevduat 946...	100	AUTO
2242	2014-01-30	16:48:00	Mehmet Şimşek: Geçmiş Günümüze Kamu...	61	AUTO
2243	2014-01-30	16:50:00	Rekabet Kurumu Kararı	0	REVER...
2245	2014-01-30	16:55:00	Türk-İtalyan İş Forumu	77	AUTO
2246	2014-01-30	16:55:00	Macanistan Merkez Bankası: Faizlen...	43	REVER...
2247	2014-01-30	17:06:00	Turizm Geliri Arh	100	AUTO
2249	2014-01-30	17:40:00	Merkez Bankasının Faiz Kararı	50	AUTO
2250	2014-01-30	17:40:00	OrtaDoğuda Laboratuvarları Kuracak	100	AUTO
2251	2014-01-30	17:50:00	Borsa Günü Yükselişle Tamamlandı	60	AUTO
2252	2014-01-30	17:54:00	"Keşke Vega Mahremiyeti Olmasa, Çikp...	95	AUTO
2253	2014-01-30	18:00:00	Ulaştırma Komut Sırtı Yüzde 27 Arttı	100	AUTO
2255	2014-01-30	18:15:00	ABD Dördüncü Çeyrekte Yüzde 3.2 Büyüdü	80	AUTO
2256	2014-01-30	18:25:00	Epdik: 2014 Yılı Doğal Gaz Tüketim...	50	AUTO
2257	2014-01-30	18:50:00	Tekrar	75	AUTO
2258	2014-01-30	20:10:00	ABD'de İstikrar: Sigortasız Talebinde...	75	REVER...
2259	2014-01-30	21:15:00	Kalkınma Bakanı: Alınur Açıklaması	60	AUTO
2260	2014-01-30	22:00:00	FED'in Kararı Ekonomileri Tutarlılasa...	40	AUTO
2261	2014-01-30	22:15:00	Bakan Yıldız: Bunu Dzelletmesi...	50	AUTO

Şekil 7 - Editör ekranı

Editör, ekran üzerinde puanlamayı ve her bir kelimeyi boyanmış olarak görmektedir. Otomatik puanlamamın yanlış olması durumunda bu ekran üzerinden puanlamayı değiştirebilmekte veya ilgisiz bir haber olduğunu düşünmesi durumunda bu sonucu kayıtlardan silebilmektedir. Böylece hatalı bulunan ve ilgisiz yakalanan haberler de kayıt altına alınmış olup sistemin başarı oranı kolaylıkla ölçülebilmektedir.

### (Editör Onayından Geçmiş Haberler İçin) Otomatik Haber Puanlama Başarı Oranları



Şekil 8 - Editör onaylı haber başarı oranları

## 2.5. Maximum Entropy

Entropy'nin kelime anlamı düzensizlik, karışıklık demektir, beklenmeyen durumların ortaya çıkma olasılığını belirtir. Bir sistemin kararsızlık derecesi yalnızca mümkün olabilecek durum sayısı ile değil, o durumlarda bulunabilme olasılıklarıyla bağlantılıdır.

$$P_i = P(X = x_i)$$

Bir (X) sisteminin bulunabileceği durumlar  $P_i$  olarak gösterildiğinde Entropy formülü aşağıdaki gibi tanımlanmaktadır:

$$H(X) = - \sum_{i=1}^n P_i \cdot \log_2 P_i$$

Logaritma tabanı çoğunlukla 2 kullanılır. Birim entropi'ye "bit" denir.

Jaynes'e göre, doğru ve tutarlı olan bilgi ile bulunan olasılıklar her sonuç için en uygun mesafedeki düzgün dağılımın elemanları olur. Olası yanlış sonuçlar için belirsizliği maksimum yapan düzgün dağılımın kullanılmasını önermektedir. Bu da entropi'nin maksimize edilmesiyle sağlanabilmektedir (Jaynes, 1957), (Baray, 2003).

$$\max H(X) = - \sum_{i=1}^n P_i \cdot \ln P_i$$

Bütün  $P_i$  değerleri eşit olduğunda  $H(X)$  maksimize edilmiş olur.  $P_i$  değerleri aşağıdaki koşulu sağlamak zorundadır:

$$\sum_{i=1}^n P_i = 1$$

Maximum Entropy, çok zor sınıflandırma işlemlerini yapmak için oldukça kullanışlı bir yöntemdir. Konuşma etiketleme, cümle tanımlama gibi konularda oldukça yaygın kullanılır. Burada önemli nokta eğitim prosedürüdür. Daha çok doğal dil işlemede kullanılmaktadır. Büyük verilerle çalışıldığında sınıflandırma işlemlerinin karmaşıklığını ve zorluğunu ortadan kaldırmaktadır.

Çalışmalarımızda geliştirdiğimiz yeni modeli kıyaslamak için bu yöntemle başvurulmuştur. Testler için SharpEntropy kodu kullanılmıştır. Kod üzerinde haber metinleri okunacak şekilde değişiklik yapılmıştır (Northedge, 2006).

Maximum Entropy yöntemi ile (veri setinin %10'luk kısmı eğitim için kullanılmıştır) aynı veriler üzerinde yapılan çalışmada GARAN kategorisi için %89.5'lük bir başarı sağlanmıştır (Nötr haberler dâhil). GENEL kategorisinde ise Nötr sonuçlar hariç tutulduğunda %99 başarı sağlanmıştır. Ancak Nötr sonuçlar hesaplama katıldığında hata oranı yükselmektedir. Editör onaylı kayıtlar içerisinde 626 adet Nötr sonucun yaklaşık %25'inde hatalı işlem yapılmaktadır. Bu da yaklaşık olarak %6'lık bir hata payına denk gelmektedir. Bu durumda GENEL kategorisi için %94'lük bir başarı sağlamaktadır.

GENEL kategorisi için Nötr kayıtlar dışarıda tutularak bakıldığında tüm kayıtlar toplamında (15.460 adet, Nötr kayıtlar sonuçların yaklaşık %23'ünü oluşturmaktadır) yeni model ile Maximum Entropy arasında %2.6'lık bir hata farkı ortaya çıkmaktadır. Ancak, hata payının hangi modele ekleneceği incelenmemiştir. Eğer Maximum Entropy'nin doğru olduğunu varsayarsak modelimizin  $\%(93-2.6)=\%90.4$  başarılı olduğu varsayımına ulaşılabilir.

Performans olarak bakıldığında her iki modelin de tarama süreleri birbirine yakın çıkmaktadır.

Yeni kurulan modelin en büyük avantajı, ilgisiz haberlerin arama dışı bırakılabilmesi olmasıdır. Maximum Entropy ve diğer öğrenen sistemler, dışarıda kalması gereken haberleri de puanlamaktadır. Yaklaşık 35.000 haberin %35'lik bölümü modelimizde sonuçlara yansıtılmamıştır. Bunun yanında, sözlük ilk tanımlanırken bilgi birikimi gerektiriyor olması ise dezavantajdır. Sözlük tanımlandıktan sonra bu dezavantaj ortadan kalkmaktadır.

### 3. Veri

Tablo 6 - Ham veri, işlenmiş veri ve sözlük bilgileri

Ham Veri (.txt)		Sözlük		İşlenmiş Veri	
Haber Grubu	Adet	Kelime Sayısı	Tarama Süresi	Puanlanan	Silinen ve Değersiz Veriler
Genel	32.519	270	11,1 dk	19.687	12.832
Şirket	2.222	457	50 sn	1.557	665

Tablo-6’da görüldüğü üzere öğrenme dönemi olarak seçilen 2014.01.01-2014.08.15 tarih aralığında toplamda 34.741 adet haber için, tanımlanmış olan sözlük yardımı ile tarama yapılarak veriler analize hazır duruma getirilmiştir.

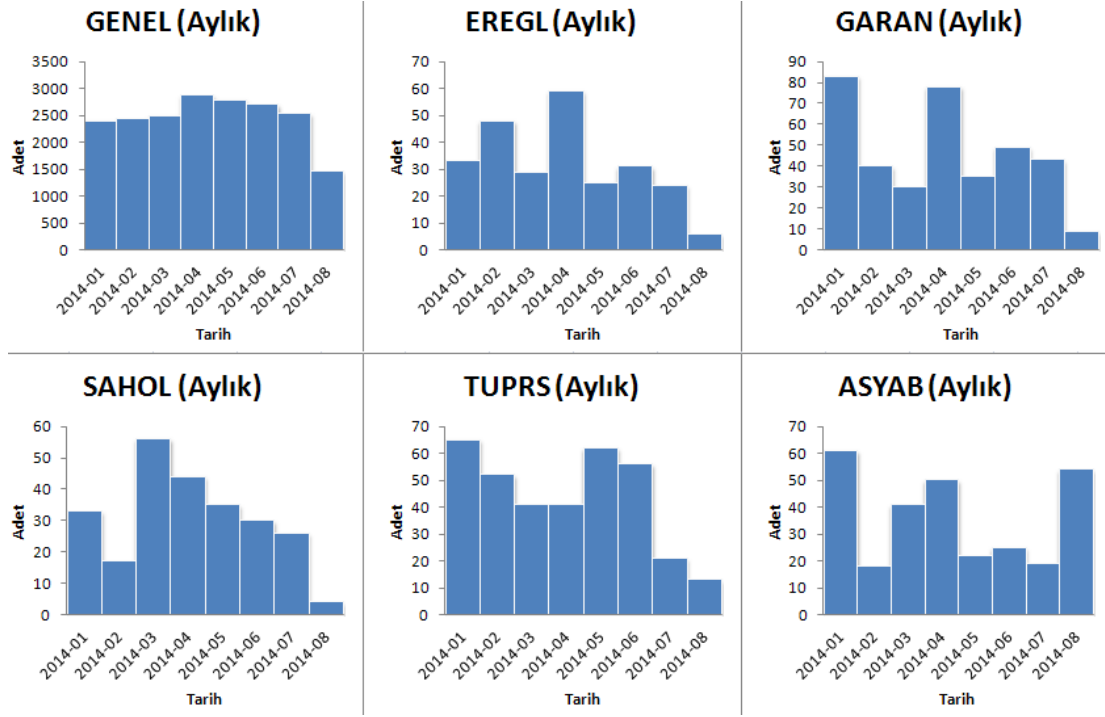
Genel haberlerin %39.5’inin tarama işlemleri ve editör onayından sonra analiz dışında bırakıldığı görülmektedir. Otomatik tarama işlemlerinden sonra 32.519 kayıttan 21.183 adet kayıt ilgili olarak işaretlenmiş ve puanlanmıştır. Editör onayında ise 1.496 adet kayıt silinmiştir.

Şirket haberlerinin ise yine aynı süreç takip edilerek %30’u analiz dışı bırakılmıştır. Tarama işlemlerinden sonra 2.222 kaydın 1.917 adedi ilgili işaretlenmiş ve editör tarafından 360 kayıt silinmiştir.

Veriler işlendikten sonra BIST100 ve şirket senetlerinin geçmiş tarihli fiyat bilgileri toplanmıştır. Haberler ve fiyat bilgileri “Günlük” ve “Saatlik” periyotlarda eşleştirilmiştir.

#### 3.1. Analiz Verisi

Tablo-6’da özet olarak gösterilmiş verilerin detayları aşağıdaki şekillerde gösterilmiştir. Grafik ve tablolar; tarama sonrasında puanlanmış ve ilgisiz haberler dışarıda bırakılmış, işlenmiş verilerden oluşmaktadır.



Şekil 9 - Aylık haber dağılımları

Üzerinde çalışılan veri 2014-01-01 ile 2014-08-15 tarih aralığında olduğu için Şekil 9'daki Ağustos ayına ait veriler diğer aylara göre düşük sayılardan oluşmaktadır.

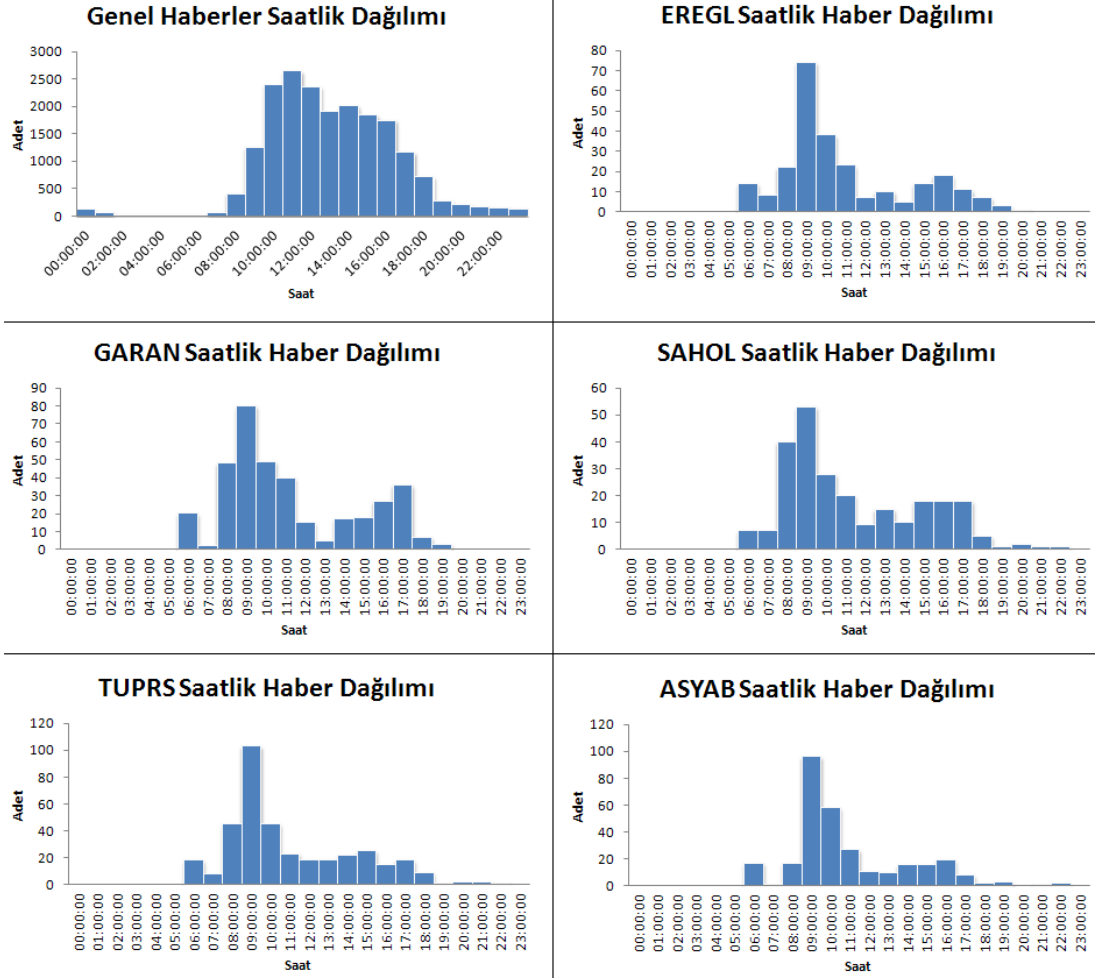
Genel kategorisindeki haber dağılımları aylık bazda birbirine oldukça yakın seyretmektedir. Buna rağmen Nisan ayında hafif bir yükselme eğiliminde olduğu göze çarpmaktadır.

Nisan ayı için haber sayısındaki artış tüm grafiklerde benzerlik göstermiştir. Nisan ayından sonra ise bir azalma söz konusudur. Aykırı tutum sergileyen grafikler genellikle siyasi etkiler altında yaşanan durumlardan kaynaklanmıştır.

ASYAB grafiği incelendiğinde Ağustos ayında (iki haftalık veri olmasına rağmen) oldukça yüksek haber akışı yaşanmıştır. Bu dönemdeki haber artışı Ziraat Bankası'nın ASYAB kurumunu satın alma konusu ve Bank Asya'nın kârındaki düşüş ile ilgilidir. Yine bu dönemde haber akışları ile paralel hisse fiyatındaki sert hareketler nedeniyle sıklıkla seans durdurma uygulanmıştır.

TUPRS için Mayıs ve Haziran döneminde, Tüpraş'a kesilen ceza ve yatırımlar ve ihracatta elde ettiği başarılar nedeniyle yoğun bir haber akışı yaşanmıştır.



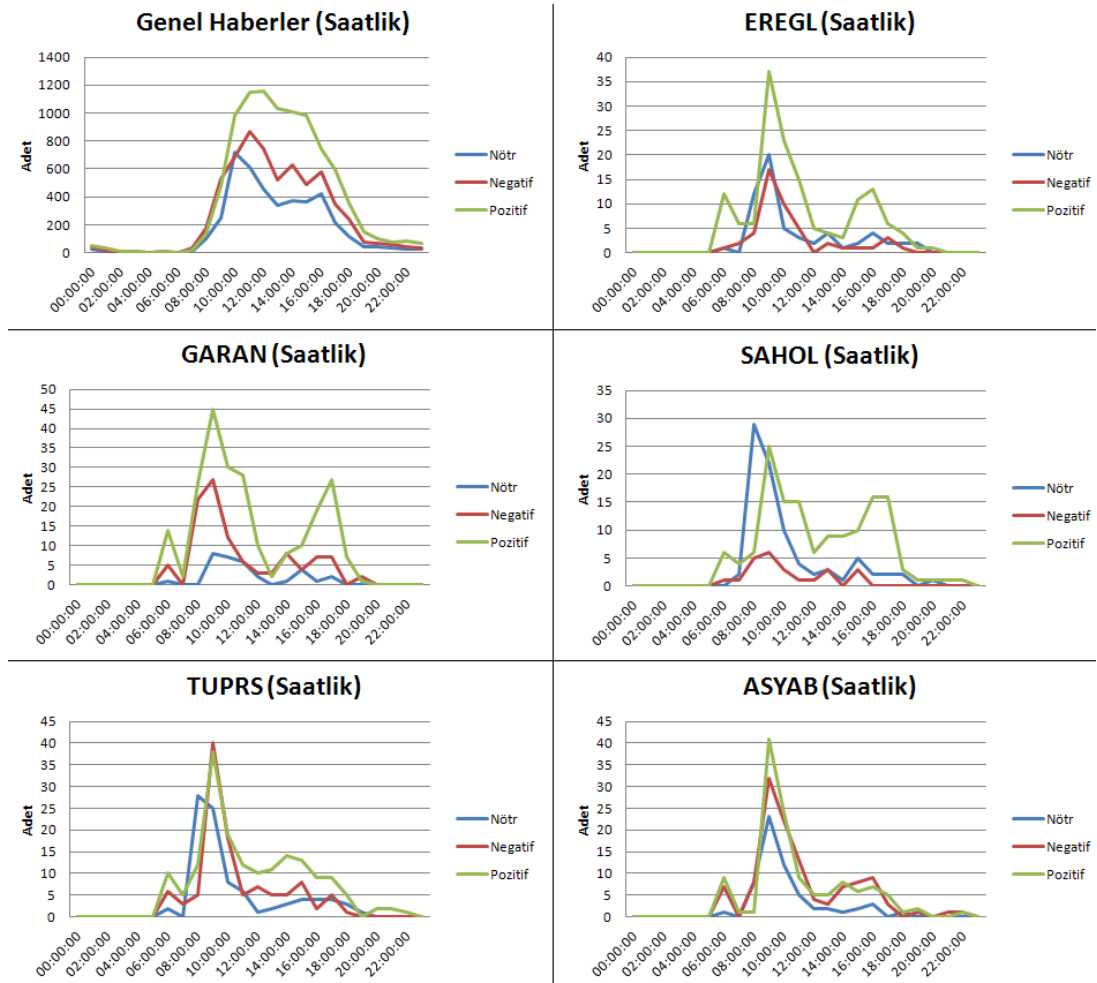


Şekil 10 - Saatlik haber dağılımları

Şirket haberleri farklı bir kaynaktan elde edildiği için haber akışları Şekil 10’da görüldüğü gibi 06:00 ile 22:00 saatleri arasında dağılmaktadır. Genel kategorisindeki haberler ise 03:00 ile 06:00 aralığında seyrekleşmekte, diğer tüm saatlerde belirginleşmektedir.

Şirket haberleri için haber akışı saat 08:00 – 11:00 ve 14:00 – 18:00 aralıklarında yoğun bir şekilde devam etmektedir, 09:00 ve 16:00’da en yüksek düzeye ulaşmaktadır. Genel haberler grafiği incelendiğinde ise 08:00 – 18:00 saatleri arasında yoğunluk görülmektedir. Bu grupta ise 11:00 ve 14:00 saatleri en yoğun bölgelerdir.

Bu veriler sadece Türkiye’de çıkan ekonomi ve şirket haberlerinden oluşmaktadır. Dolayısıyla yurtdışı piyasalarındaki haber akışları ile bir ilişkisi bulunmamaktadır.



Şekil 11 - Saatlik haber dağılımları (haberlerin eğilimine göre)

Haberlerin, eğilime göre incelenmesinin amacı pozitif ve negatif haberlerin hangi saat diliminde yoğunlaştığını tespit etmek ve bu aralıklardaki fiyat hareketlerini incelemektir. Burada “Haberler mi fiyat hareketlerinden etkileniyor yoksa fiyatlar mı haber akışından etkileniyor?” sorusunun cevabı aranmaktadır.

Genel haberler (saatlik) grafiği incelendiğinde (Şekil 11) saat 07:00 ve 09:00 aralığında negatif haber sayısının pozitiften fazla olduğu dikkat çekmektedir. Buradan negatif haberlerin bu saat aralığında yayınlandığı ve buna göre alım satım emirlerinin saat 10.00’den sonra yapılmasının mantıklı olabileceği düşünülebilir. Tablo 7 üzerinde detaylı veriler incelenebilir.

EREGL ve SAHOL verileri için negatif ağırlıklı saat dilimi bulunmamaktadır. TURPS için saat 09:00 diliminde negatif haberler fazla çıkmıştır. GARAN için saat 13:00 dilimi ve 19:00’da negatif haber akışı tespit edilmiştir.

ASYAB verilerinin farklı bir karakteristiğe sahip olduğu düşünülmektedir. 15:00 ve 16:00'da gelen negatif yönlü haberlerin fiyatlardaki tepki hareketlerini baskılama yönünde olabileceği varsayılabilir. Nitekim bu saatlerde çoğunlukla eksi getiri görülmektedir, ancak, sıklıkla seans kapama uygulandığından dolayı bu varsayımı doğrulamak için veriler yetersiz kalmaktadır. Saatlik detaylar için Tablo 8 incelenebilir.

Tablo 7 - Genel haber saatlik dağılımlar (eğilime göre)

Genel Haber Saatlik Dağılım (eğilime göre)							
Saat	Negatif	Nötr	Pozitif	Toplam	Negatif/Toplam	Nötr/Toplam	Pozitif/Toplam
00:00:00	41	25	47	<b>113</b>	0.36	0.22	0.42
01:00:00	18	12	33	<b>63</b>	0.29	0.19	0.52
02:00:00	3	4	10	<b>17</b>	0.18	0.24	0.59
03:00:00	6	2	11	<b>19</b>	0.32	0.11	0.58
04:00:00	2	2	5	<b>9</b>	0.22	0.22	0.56
05:00:00	8	2	9	<b>19</b>	0.42	0.11	0.47
06:00:00	1	2	1	<b>4</b>	0.25	0.50	0.25
<b>07:00:00</b>	<b>31</b>	<b>11</b>	<b>20</b>	<b>62</b>	<b>0.50</b>	<b>0.18</b>	<b>0.32</b>
<b>08:00:00</b>	<b>174</b>	<b>97</b>	<b>132</b>	<b>403</b>	<b>0.43</b>	<b>0.24</b>	<b>0.33</b>
<b>09:00:00</b>	<b>531</b>	<b>247</b>	<b>479</b>	<b>1257</b>	<b>0.42</b>	<b>0.20</b>	<b>0.38</b>
10:00:00	690	716	984	<b>2390</b>	0.29	0.30	0.41
11:00:00	872	616	1152	<b>2640</b>	0.33	0.23	0.44
12:00:00	745	456	1155	<b>2356</b>	0.32	0.19	0.49
13:00:00	521	343	1037	<b>1901</b>	0.27	0.18	0.55
14:00:00	631	372	1008	<b>2011</b>	0.31	0.18	0.50
15:00:00	485	368	985	<b>1838</b>	0.26	0.20	0.54
16:00:00	582	420	741	<b>1743</b>	0.33	0.24	0.43
17:00:00	344	217	596	<b>1157</b>	0.30	0.19	0.52
18:00:00	239	118	357	<b>714</b>	0.33	0.17	0.50
19:00:00	76	46	150	<b>272</b>	0.28	0.17	0.55
20:00:00	67	42	98	<b>207</b>	0.32	0.20	0.47
21:00:00	58	35	74	<b>167</b>	0.35	0.21	0.44
22:00:00	44	25	83	<b>152</b>	0.29	0.16	0.55
23:00:00	38	24	70	<b>132</b>	0.29	0.18	0.53

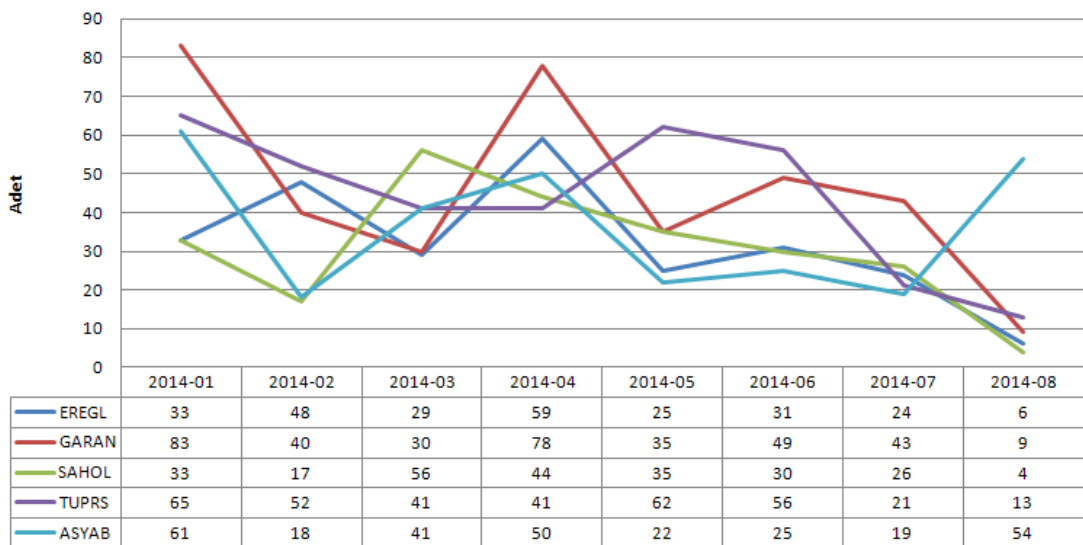
(Tablo 7 ve Tablo 8'de yer alan Negatif, Nötr, Pozitif ve Toplam sütunlarının birimleri Adet cinsindedir)

Tablo 8 - Asyab Saatlik Haber Dağılımı (eğilime göre)

ASYAB Saatlik Haber Dağılım							
Saat	Negatif	Nötr	Pozitif	Toplam	Negatif/Toplam	Nötr/Toplam	Pozitif/Toplam
06:00:00	7	1	9	17	0.41	0.06	0.53
07:00:00	0	0	1	1	0.00	0.00	1.00
<b>08:00:00</b>	<b>8</b>	<b>8</b>	<b>1</b>	<b>17</b>	<b>0.47</b>	<b>0.47</b>	<b>0.06</b>
09:00:00	32	23	41	96	0.33	0.24	0.43
10:00:00	22	12	24	58	0.38	0.21	0.41
<b>11:00:00</b>	<b>13</b>	<b>5</b>	<b>9</b>	<b>27</b>	<b>0.48</b>	<b>0.19</b>	<b>0.33</b>
12:00:00	4	2	5	11	0.36	0.18	0.45
13:00:00	3	2	5	10	0.30	0.20	0.50
14:00:00	7	1	8	16	0.44	0.06	0.50
<b>15:00:00</b>	<b>8</b>	<b>2</b>	<b>6</b>	<b>16</b>	<b>0.50</b>	<b>0.13</b>	<b>0.38</b>
<b>16:00:00</b>	<b>9</b>	<b>3</b>	<b>7</b>	<b>19</b>	<b>0.47</b>	<b>0.16</b>	<b>0.37</b>
17:00:00	3	0	5	8	0.38	0.00	0.63
18:00:00	0	1	1	2	0.00	0.50	0.50
19:00:00	1	0	2	3	0.33	0.00	0.67
20:00:00	0	0	0	0	0.00	0.00	0.00
21:00:00	1	0	0	1	1.00	0.00	0.00
22:00:00	1	0	1	2	0.50	0.00	0.50
23:00:00	0	0	0	0	0.00	0.00	0.00

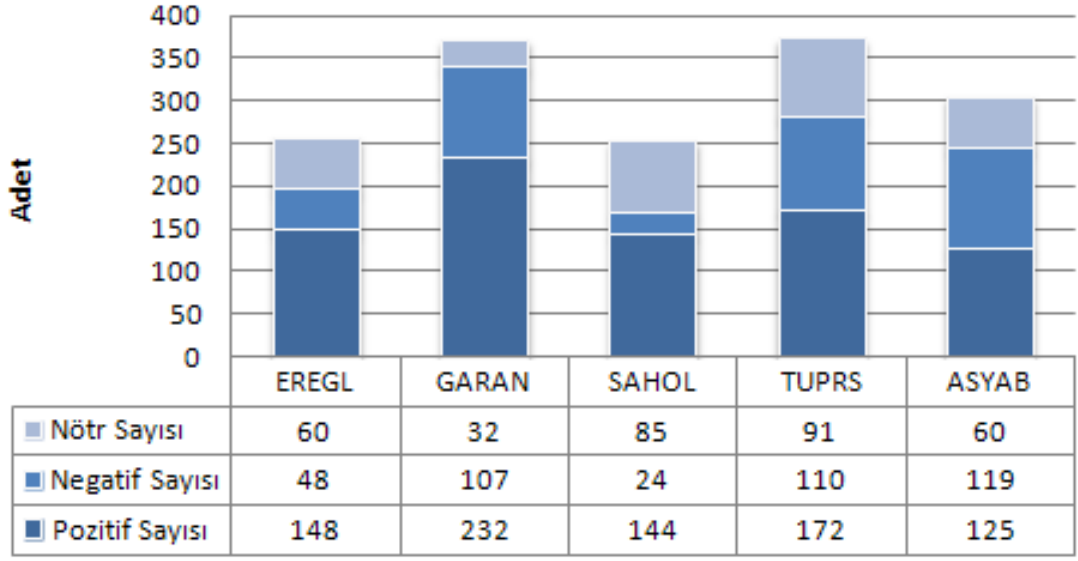
(Tablo 7 ve Tablo 8’de koyu renkli gösterilen satırlar negatif haberlerin ağırlıklı olduğu saatleri belirtmektedir)

### Şirket Haberleri (Aylık)



Şekil 12 - Şirket haberleri için karşılaştırmalı aylık dağılımlar

## Toplam Şirket Haber Dağılımı



Şekil 13 - Şirket haberlerinin eğilimlerine göre aylık karşılaştırmalı dağılımı

Şekil 12 ve Şekil 13'te sadece şirketler için (Genel kategorisi dışarıda bırakılarak) aylık dağılımlar karşılaştırmalı olarak incelenmiştir.

Şekil 12'de TUPRS grafiğinin diğerlerinden ayrıştığı görülmektedir. ASYAB ise Ağustos ayında farklılık göstermektedir.

Şekil 13 incelendiğinde, pozitif haber sayısının ASYAB dışında negatif haber sayısının en az iki katı olduğu görülmektedir. Genel itibari ile pozitif ağırlıklı haber yayınlanmıştır. Nötr haber sayısının da SAHOL ve EREGL için negatif haber sayısından yüksek olduğu göze çarpmaktadır.

### 3.2. Analiz

Bölüm 3.1’de detaylı olarak ele alınan veriler, istatistiksel olarak anlamlı hale getirilerek günlük (Tablo 9) ve saatlik (Tablo 10) olacak şekilde, genel haberler ve her bir şirket için ayrı ayrı hazırlanmıştır. Şirket tablolarının örnekleri eklenmemiştir, sonuçlara ve grafiklere ileriki sayfalarda değinilecektir.

#### Tablo 9 ve Tablo 10’da gösterilen sütunlar

**Tarih:** Günlük periyotta gruplanmıştır, hafta sonları ve resmî tatiller veriler içerisinden çıkartılmıştır. Tatil günlerine denk gelen haberler, ilk iş gününe aktarılmıştır.

**Saat:** Saatlik periyotta gruplandırılmıştır. Hisse senetleri piyasasının açılış saatinden öncesi saat 09:00 dilimine eklenmiştir. Kapanış saatinden sonraki kayıtlar ise bir sonraki günün saat 09:00 dilimine aktarılmaktadır. Yani saatlik çalışma verileri 09:00 ile 17:00 aralığındadır.

**BIST100 Kapanış:** Borsa İstanbul 100 Endeksi kapanış fiyatı.

**BIST100 Getiri:** Bugünün (bu saatin) kapanış fiyatının bir önceki günün (saatin) kapanış fiyatına oranı ile hesaplanır.

$$Getiri = \left( \frac{\text{kapanış}_{(t)}}{\text{kapanış}_{(t-1)}} \right) - 1$$

**Haber Binary:** Günlük veya saatlik periyotta gruplanan haberlerin eğilime göre sayısal toplamları alınarak hesaplanır (pozitif = 1, negatif = -1 veya nötr = 0). Tek değer belirtir.

**Haber Avg. Ratio:** 2.4. Tarama ve Sınıflandırma başlığı altında verilen pozitifPuan ve negatifPuan formülleri ile her bir haber için hesaplanan değerlerin ilgili zaman dilimindeki aritmetik ortalamasıdır. Formüllerde hesaplanan değerler [0-100] aralığındadır, ancak Tablo 9 ve Tablo 10’da aritmetik ortalaması alınan değerler [100-0-(-100)] aralığındadır. Formülde hesaplanan Nötr = 50 değerini alırken, Ratio hesaplanırken 0 “Sıfır” değerini almaktadır.

Tablo 9 - Genel haber kategorisi için Günlük periyot çalışma verisi

Tarih	BIST100 Kapanış	BIST100 Getiri	Haber Binary	Haber Avg. Ratio
02.01.2014	66985.82	0.000	1	12.54
03.01.2014	65967.14	-0.015	1	12.87
06.01.2014	68021.99	0.031	1	7.45
07.01.2014	68598.61	0.008	1	8.87
08.01.2014	67331.7	-0.018	1	15.19
09.01.2014	66413.59	-0.014	-1	-2.02
10.01.2014	67912.11	0.023	1	15.00
13.01.2014	68062.29	0.002	1	28.78
14.01.2014	68072.51	0.000	1	21.98

Tablo 10 - Genel haber kategorisi için Saatlik periyot çalışma verisi

Tarih	Saat	BIST100 Kapanış	BIST100 Getiri	Haber Binary	Haber Avg. Ratio
02.01.2014	09:00:00	66375.20	0.000	1	24.56
02.01.2014	10:00:00	66891.36	0.008	1	65.00
02.01.2014	11:00:00	66896.48	0.000	-1	-2.00
02.01.2014	12:00:00	66510.72	-0.006	-1	4.80
02.01.2014	14:00:00	66056.93	-0.007	1	4.93
02.01.2014	15:00:00	66076.54	0.000	-1	2.32
02.01.2014	16:00:00	66257.44	0.003	1	40.00
02.01.2014	17:00:00	66985.76	0.011	1	-0.50
03.01.2014	09:00:00	66853.60	-0.002	-1	-12.25

Tablo 9 ve Tablo 10'da tüm kayıtlar listelenmemiştir. Günlük periyotta 160 günlük, saatlik periyotta da 1240 saatlik gruplanmış veri ile çalışılmıştır.

Getiri ile haber istatistiksel verileri arasında korelasyona (ilişki katsayısı) bakılmıştır. Hesaplamalar için Excel uygulaması kullanılmıştır. Tablo 13'de HaberBinary, HaberAvgRatio ve HaberAdet ile Bist100Getiri sütunlarının, günlük ve saatlik periyotlar için korelasyonları bulunmuştur. Bu hesaplamalarda haberlerin gecikmeli olarak etkili olabileceği düşünülmüştür. Haberlerin günlük olarak 10 güne kadar, saatlik olarak da 10 saate kadar ötelenerek ilgili zamandaki BIST100 getirisi ile korelasyonu da hesaplanmıştır.

**Korelasyon:** İki rassal değişken arasında doğrusal ilişki olup olmadığını ve büyüklüğünü gösterir.

$$Correl(X, Y) = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

$\bar{x}, \bar{y} \rightarrow Average(dizi1), Average(dizi2)$

Korelasyona “r” dersek alabileceği değerler -1 ve +1 aralığında olabilecektir.

$$-1 \leq r \leq +1$$

İlişki katsayısı 0 “sıfır” ise karşılaştırılan değerler arasında ilişki yoktur. -1 veya +1 değerlerini alıyorsa tam doğrusal bir ilişki vardır. Pozitif “r” direkt yönlü bir doğrusal ilişki, negatif “r” ise ters yönlü bir doğrusal ilişki bulunduğunu gösterir (Kumar & Chaudhary, 2004).

Tablo 13’de hesaplanmış olan korelasyon sonuçları gösteriyor ki; haberler (binary, avg., adet) ile BIST100 getiri arasında doğrusal bir ilişki bulunmamaktadır. Bu nedenle daha karmaşık modellere ihtiyaç duyulmaktadır.

Tablo 11 - Korelasyon Günlük Karşılaştırma

<b>Günlük Periyot</b>				
	Toplamda İlişki Yönü	En Yüksek Değere Göre İlişki Yönü	Direk Yönlü Max. Değer	Ters Yönlü Max. Değer
HaberBinary-BistGetiri	+	-	Gecikmesiz	1 Gün
HaberAvgRatio-BistGetiri	-	-	7 Gün	1 Gün
HaberAdet-BistGetiri	+	+	4 Gün	10 Gün

Tablo 12 - Korelasyon Saatlik Karşılaştırma

<b>Saatlik Periyot</b>				
	Toplamda İlişki Yönü	En Yüksek Değere Göre İlişki Yönü	Direk Yönlü Max. Değer	Ters Yönlü Max. Değer
HaberBinary-BistGetiri	-	-	Gecikmesiz	10 Saat
HaberAvgRatio-BistGetiri	-	-	9 Saat	10 Saat
HaberAdet-BistGetiri	-	+	8 Saat	7 Saat

**Günlük periyotta** (Tablo 11), toplamda ilişki yönü pozitif, aldığı değere göre negatif çıkmıştır. Sadece HaberBinary “Gecikmesiz” veride en yüksek değeri almıştır.

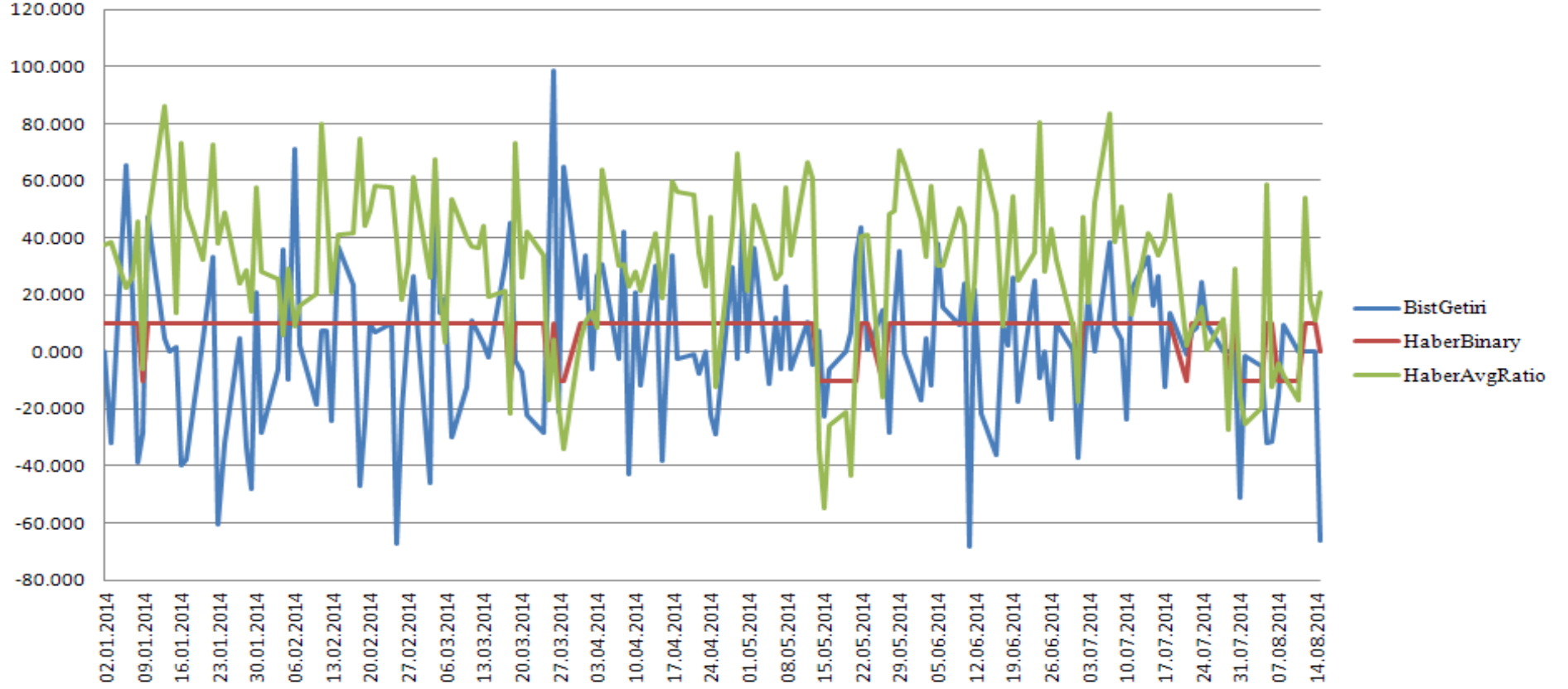
**Saatlik periyotta** (Tablo 12), genel itibari ile negatif yönlü ilişki çıkmıştır. En yüksek değerler, yaklaşık 1 Gün’e denk gelen 7-10 saat dilimi içerisinde oluşmuştur.



Tablo 13 - Korelasyon Detay (Gecikmeli Haber ve Bist100 Getiri)

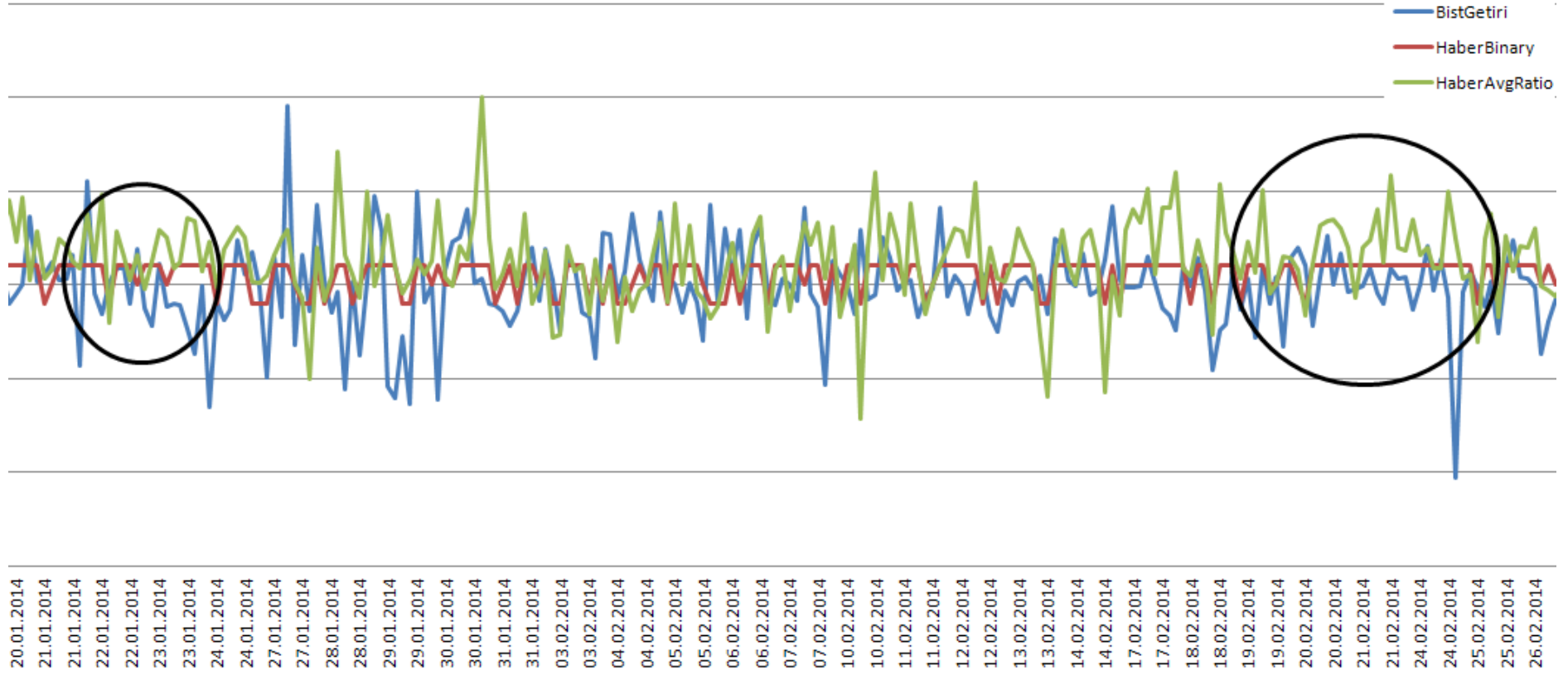
<b>Günlük Gecikmeli (HaberBinary-BistGetiri)</b>											
<b>Gecikme</b>	Yok	1 Günlük	2 Günlük	3 Günlük	4 Günlük	5 Günlük	6 Günlük	7 Günlük	8 Günlük	9 Günlük	10 Günlük
<b>Korelasyon</b>	0.0340	-0.1221	0.0124	-0.0767	0.0087	-0.0062	-0.0291	0.0007	-0.0854	0.0137	0.0222
<b>Günlük Gecikmeli (HaberAvgRatio-BistGetiri)</b>											
<b>Gecikme</b>	Yok	1 Günlük	2 Günlük	3 Günlük	4 Günlük	5 Günlük	6 Günlük	7 Günlük	8 Günlük	9 Günlük	10 Günlük
<b>Korelasyon</b>	0.0130	-0.1635	-0.0401	-0.1159	-0.0268	-0.1090	-0.0072	0.1189	-0.0833	0.0180	0.0264
<b>Günlük Gecikmeli (HaberAdet-BistGetiri)</b>											
<b>Gecikme</b>	Yok	1 Günlük	2 Günlük	3 Günlük	4 Günlük	5 Günlük	6 Günlük	7 Günlük	8 Günlük	9 Günlük	10 Günlük
<b>Korelasyon</b>	0.0697	-0.0357	0.0197	0.0522	-0.0842	0.0280	0.0309	0.0619	-0.0014	-0.0574	0.1000
<b>Saatlik Gecikmeli (HaberBinary-BistGetiri)</b>											
<b>Gecikme</b>	Yok	1 Saatlik	2 Saatlik	3 Saatlik	4 Saatlik	5 Saatlik	6 Saatlik	7 Saatlik	8 Saatlik	9 Saatlik	10 Saatlik
<b>Korelasyon</b>	0.0210	-0.0941	0.0165	-0.0793	0.0216	-0.0275	-0.0552	0.0180	-0.0599	0.0601	-0.0069
<b>Saatlik Gecikmeli (HaberAvgRatio-BistGetiri)</b>											
<b>Gecikme</b>	Yok	1 Saatlik	2 Saatlik	3 Saatlik	4 Saatlik	5 Saatlik	6 Saatlik	7 Saatlik	8 Saatlik	9 Saatlik	10 Saatlik
<b>Korelasyon</b>	0.0351	-0.1538	-0.0545	-0.1022	0.0079	-0.0989	-0.0342	0.1389	-0.0241	0.0690	0.0152
<b>Saatlik Gecikmeli (HaberAdet-BistGetiri)</b>											
<b>Gecikme</b>	Yok	1 Saatlik	2 Saatlik	3 Saatlik	4 Saatlik	5 Saatlik	6 Saatlik	7 Saatlik	8 Saatlik	9 Saatlik	10 Saatlik
<b>Korelasyon</b>	0.0177	-0.0577	0.0090	0.0599	-0.0866	-0.0271	-0.0113	0.0107	-0.0399	-0.1158	0.0827

### Genel Haberler - BIST100 Getiri



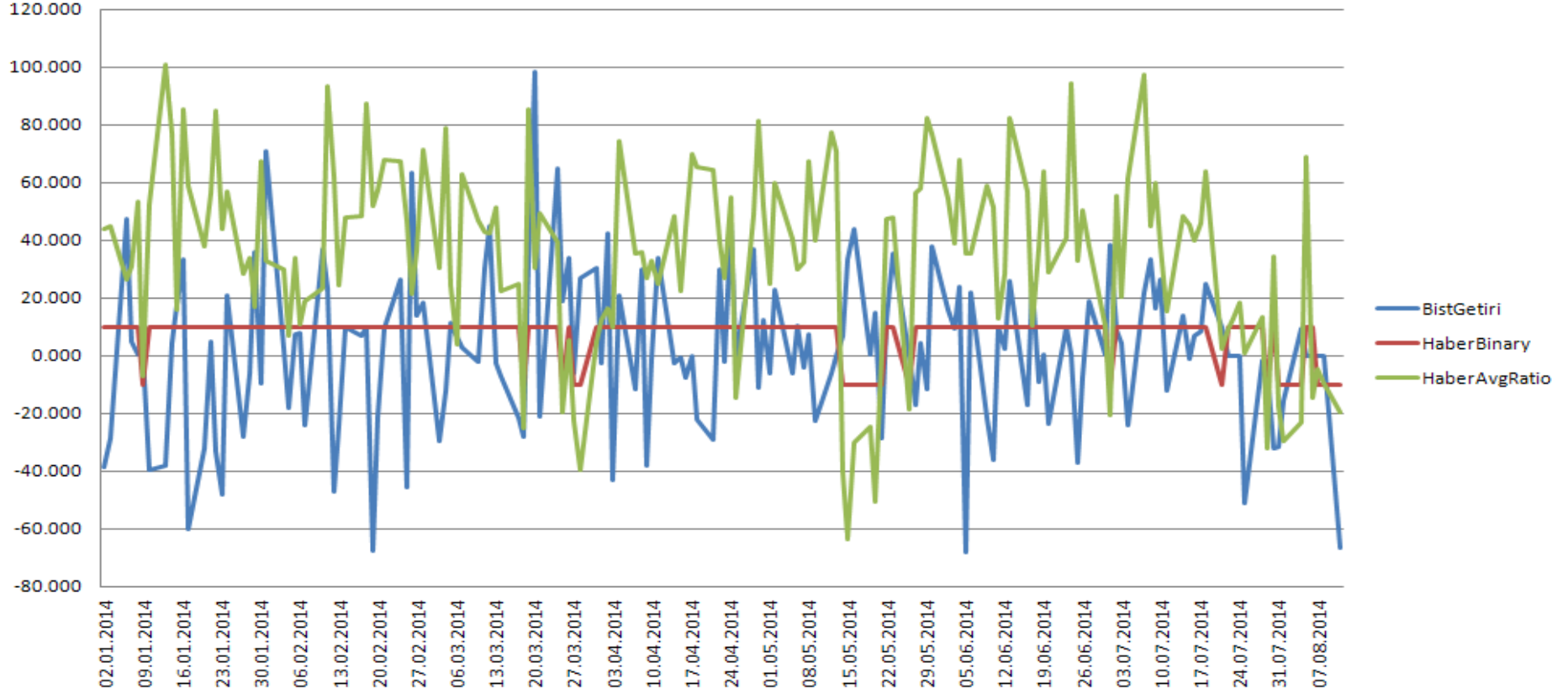
Şekil 14 - Genel Haberler Bist100 Getiri (Günlük) Karşılaştırması

### Genel Haberler - BIST100 Getiri (Saatlik)

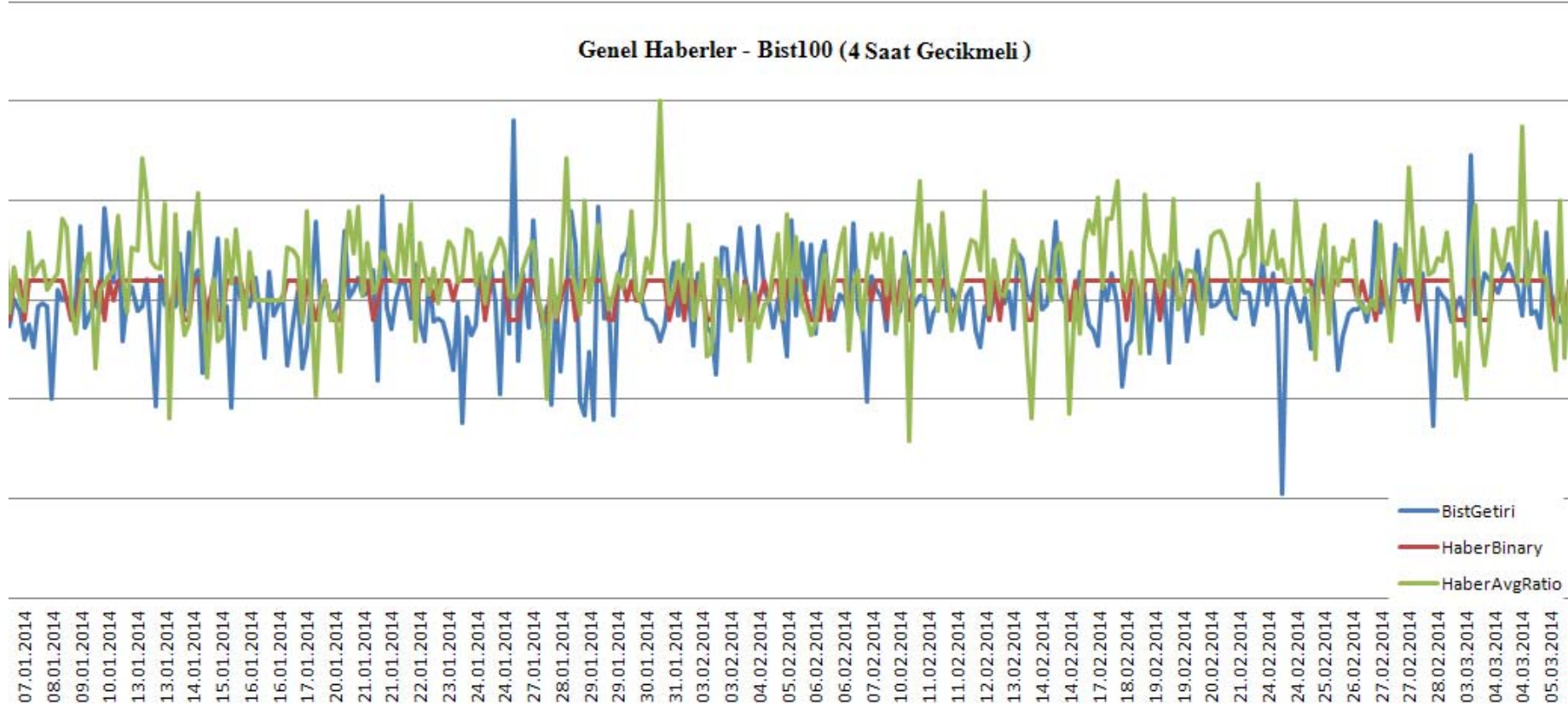


Şekil 15 - Genel Haberler Bist100 Getiri (Saatlik) Karşılaştırması

Genel Haberler - Bist100 Getiri (4 Gün Gecikmeli)



Şekil 16 - 4 Gün Gecikmeli Genel Haberler - Bist100 Getiri (Günlük)



Şekil 17 - 4 Saat Gecikmeli Genel Haberler - Bist100 Getiri (Saatlik)

Şekil 14, Şekil 15, Şekil 16 ve Şekil 17’de kullanılan HaberBinary, HaberAvgRatio ve BistGetiri veri setleri, grafiklerde belirgin gösterilmek için belli katsayılar ile çarpılmıştır.

Şekil 14’de, günlük periyot için genel haberler ile BIST100 getirileri karşılaştırılmıştır. Grafik üzerinde, gözle görülebilecek çok belirgin bir desen bulunmamaktadır. 15.05.2014-22.05.2014 tarih aralığında benzer hareket çıkmıştır. Kimi zaman Bist100, kimi zaman da haberler öncü hareket yapmış görünmektedir.

Şekil 15’de, saatlik periyot için genel haberler ile BIST100 getirileri karşılaştırılmıştır. Veri setinde 1.240 kayıt bulunduğu için sadece belli bir kesiti şekilde gösterilebilmiştir. Genel itibari ile günlük grafiğe benzer karakteristik gösterirken, HaberBinary’nin hareketi göze çarpmıştır. Siyah daireler içerisine alınan bölgelerde, belli bir süre pozitif haber yayınlandıktan sonra ani düşüşler olduğu düşünülmektedir. Yine aynı şekilde tersinin olup olmadığı merak edilmektedir. Bu varsayımımızın doğruluğunu gözlemleyebilmek için bir Portföy uygulaması yazılmıştır. Detayları ve sonuçları bölüm 3.3’de anlatılmıştır.

Şekil 16’da, genel haberler ile BIST100 getirileri (4 gün gecikmeli) karşılaştırılmıştır. Tablo 13 ile birlikte incelendiğinde, ters yönlü ilişkinin bir miktar arttığı görülmektedir. Bu değişim, korelasyona bakılmadan da gözle farkedilebilmektedir.

Şekil 17’de, genel haberler ile BIST100 getirileri (4 saat gecikmeli) karşılaştırılmıştır. Burada da Şekil 15’de bahsedilen HaberBinary verisi, 4 saatlik gecikme alınmasına rağmen belirginliğini kaybetmemiştir.

Şirket haberleri için, grafik ve tablolar eklenmemiştir. Aynı çalışma şirket haberlerine uygulandığında benzer sonuçlar alınmaktadır.

ASYAB için korelasyon hesaplandığında ilginç şekilde diğer şirket ve genel haber kategorisinden oldukça yüksek ilişki çıktığı görülmüştür. Ancak, yapılan incelemede seans kapama nedeniyle, getirilerin çoğunlukla “sıfır” hesaplanmasının buna neden olabileceği düşünülmektedir.

### 3.3. Portföy

Şekil 15’de gösterilen HaberBinary ile BistGetiri verileri arasında korelasyon olmamasına rağmen, belirli bir süre aynı çizgide devam eden haber akışının (pozitif veya negatif) ardından bazı noktalarda getirilerde büyük bir değişim yaşandığı gözle tespit edilmiştir. Tespit edilen desenin bir sürekliliğinin olup olmadığı konusu araştırmalarımızı farklı bir noktaya taşımıştır: Daha karmaşık bir model ile haber akışlarından yararlanılabilir mi? Bu konuda zaman ölçekli portföy üzerine yapılmış çalışmalar mevcuttur (Chen & Szeto, 2012).

**Hipotez:** “Belirli bir süre aynı çizgide devam eden haber akışının (pozitif veya negatif) ardından getirilerde ani bir değişim yaşanmaktadır.”

Bu hipotezi doğrulayabilmek ve uygun parametreleri kestirebilmek için bir Portföy Stratejisi uygulaması (Şekil 18) geliştirilmesine karar verilmiştir. Uygulama, girilen parametreler ile geçmiş veri üzerinde ilgili deseni arayarak sonuca varacaktır.

İkinci bir portföy uygulaması ise yatırımcıya en iyi getiriye sağlayacak şekilde (getiri odaklı) tasarlanmıştır.

Uygulamaya, ilgili yatırım aracının tarihsel işlenmiş verileri aktararak, çalışılacak ortam hazırlanır. Nötr veriler hesaplama dışı bırakılmıştır.

Giriş parametreleri aşağıda tanımlanmıştır:

$$k_{(k \in \mathbb{N}^+)} = \{3, 4, 5, 6, 7, 8, 9, 10, \dots\}$$

$$d_{(d \in \mathbb{R})} = \{1.5, 2, 2.5, 3, 4, 5, \dots\}$$

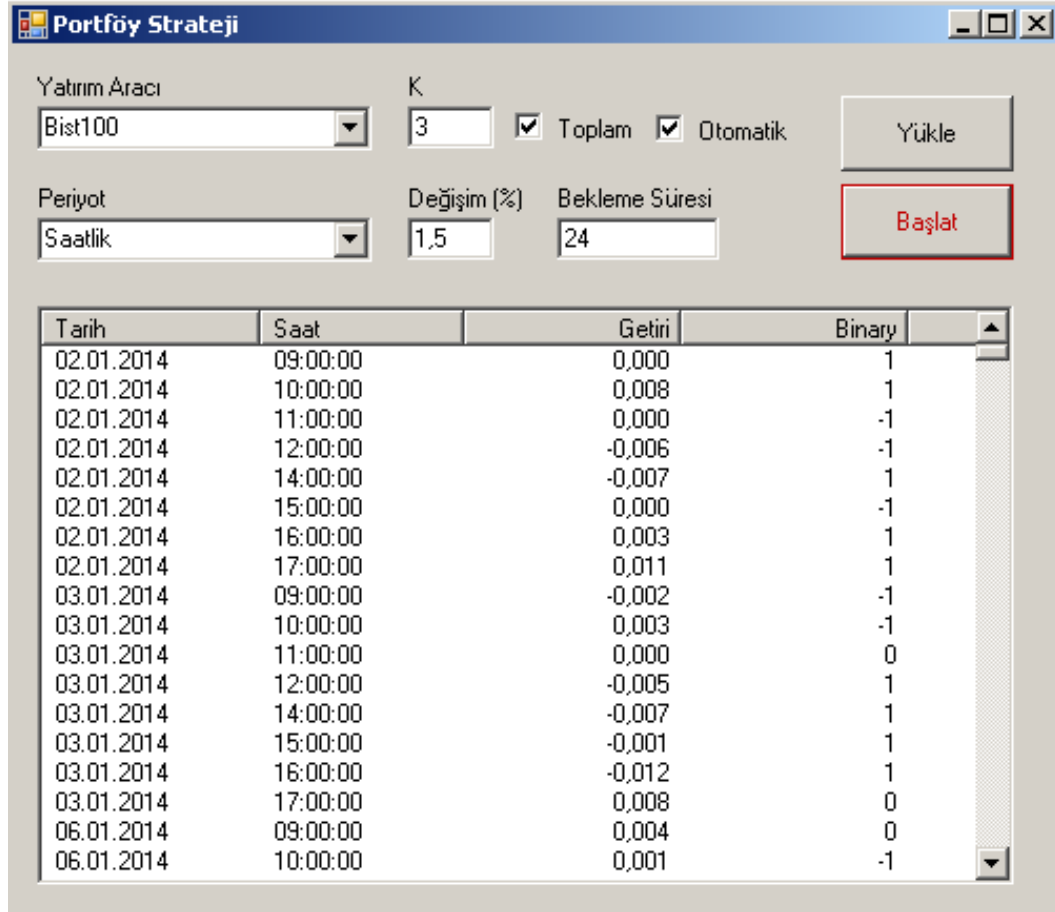
**k:** aynı değerin tekrar sayısı (HaberBinary),

**d:** % değişim (BistGetiri),

**y:** yöntem (toplam/tekil)

**b:** bekleme süresi (varsayılan maksimum değerler: günlük = 7 gün, saatlik = 24 saat)

y parametresi tekil seçildiğinde uygulama %d değişiminin tek seferde aldığı değeri kontrol eder. Toplam seçeneği tercih edildiğinde ise değişimlerin (pozitif/negatif) belirtilen değere toplamda ilk ulaştığı noktaya kadar bekler.



Şekil 18 - Portföy Stratejisi Uygulaması

**Portföy-I çalışma prensibi:** Günlük veya saatlik veriler üzerinde, giriş parametrelerinde belirtilen tüm elemanlar için: k adet tekrar eden desen yakalandığında, verilen bekleme süresi içerisinde %d'lik değişime ulaşırsa al/sat (yeni portföy oluşturma), yine bekleme süresi içerisinde tek seferde veya toplamda %d değerine ulaştığında al/sat işlemi (portföyü elden çıkartma) yapılır. Eğer bekleme süresi dolduğunda %d değerine ulaşılamaz ise kâr veya zarar durumuna bakılmaksızın portföy elden çıkarılır. Sonuçta en yüksek getiri değerine ulaşan strateji tercih edilerek yatırım planı uygulanır. Şekil 22'de akış şeması gösterilmiştir. (Portföy I - sol)

Şirket haberlerinin saatlik periyotta yeterli verisi olmadığı gerekçesiyle, uygulamada sadece günlük periyot değerlendirmeye alınmıştır.

Genel haber kategorisi için Tablo 14'de saatlik ve Tablo 15'de günlük periyotlarda:

$$k = [3-6],$$

$$d = [1.5-4]$$



aralıklarında eğitim dönemi sonuçları görülmektedir. Pozitif (+) ve Negatif (-) Binary değerleri için; seçilen k ve d değerlerinde kâr edilen işlemlerin adetleri, iptal işlemlerinin adetleri (bekleme süresi dolduğu için vazgeçilen portföy), getirileri ve toplam getirileri hesaplanmıştır.

Saatlik periyot (tekil) ve günlük periyot (tekil) için pozitif getiriler elde edilmiş, toplam yöntemi için negatif getiriler ağırlık kazanmıştır. Bu durumda hipotezimiz tekil yöntem için doğru görünmektedir. Tekil yöntemde ani hareketler tespit edilebilmektedir, ancak; getirisi düşük kalmaktadır. Test girdilerimizden  $d = 1.5$ , BIST100 fiyat değişimlerinde en yüksek tek seferlik değişimi göstermektedir.

Günlük periyotta bu değişimler saatlik periyoda göre daha çok yakalanmış, ancak; saatlik periyottaki getiriler daha yüksek elde edilmiştir. Negatif haberlerde sonuç beklenenin aksine farklı çıkmıştır. Burada getirinin yukarı çıktıktan sonra aşağı inmediği, tekrar yukarı hareketini devam ettirdiği görülmüştür. Bu da bir bakıma kötü haber karşısında tepki/fırsat alımı olarak düşünülebilir. Eğer negatif haberler için strateji bu yönde değiştirilir ise eksi getiriler artı yönde değişmiş olacaktır.

**Saatlik periyot:** Tablo 14 incelendiğinde, (tekil yöntem) tüm k değerleri için en yüksek getiri  $d = 1.5$  için çıkmıştır; yani seçilen test aralıklarında BIST100’de, tek seferdeki en yüksek değişim  $1.5$ ’tur. Getiri olarak bakıldığında da yine aynı değerler için maksimum  $15.5$ ’lik bir kazanç sağlamaktadır.

Yine aynı tabloda, toplam yönteminin sonuçlarına bakıldığında stratejimiz  $k = 5$ ,  $d = 2$  için en iyi getiriye sağlamış, pozitif binary için zarar etmiştir.

**Günlük periyot:** Tablo 15 incelendiğinde, (tekil yöntem) tüm k değerleri için en yüksek getiri  $d = 1.5$  ve  $d = 2$  için çıkmıştır. Getiri olarak bakıldığında  $d = 1.5$  daha iyi sonuç vermiştir ve  $4.3$  getiri sağlamıştır.

Yine aynı tabloda, toplam yönteminin sonuçlarına bakıldığında;  $k = 3$ ,  $d = 2$  için en iyi getiri sağlanmış, geri kalan tüm değerlerde zarar oluşmuştur.

Bu durumda, tekil yöntem en iyi seçenek olarak görünmekle beraber, toplam yöntemde negatif binary için tepki stratejisi uygulanarak daha iyi getiri elde edilmesi mümkün görünmektedir. Yine bekleme süresi artırılarak etki analizi yapılabilir.

Tablo 14 - Saatlik Portföy Stratejileri (Genel Haberler/Bist100)

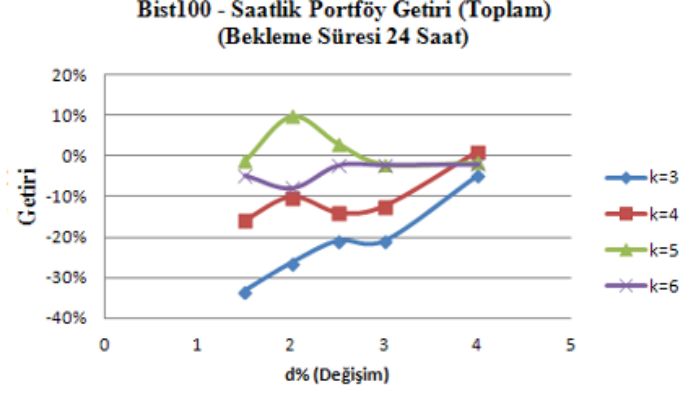
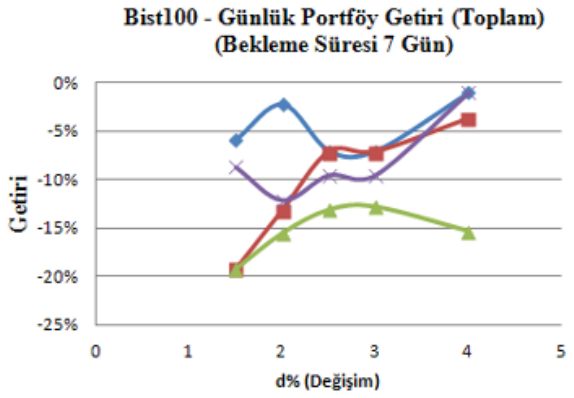
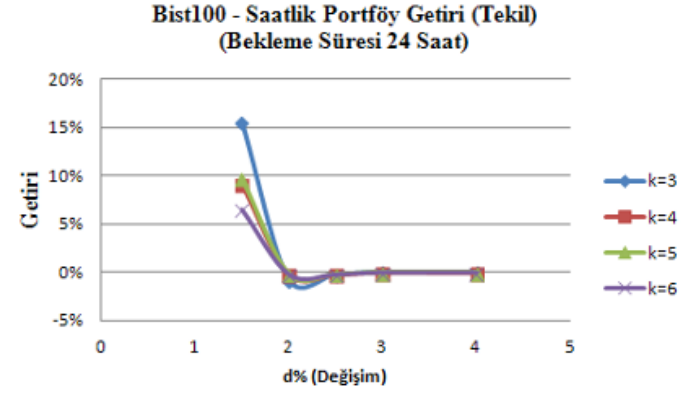
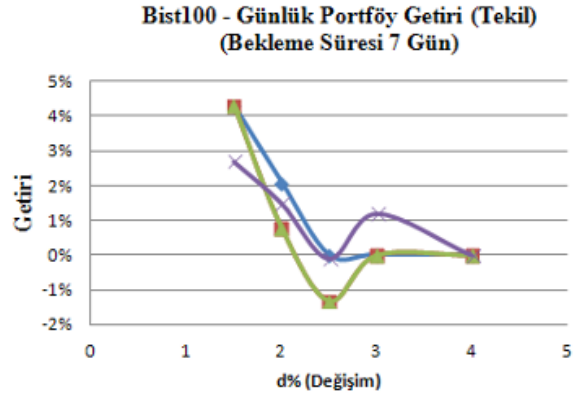
Saatlik (Tekil) - Bekleme Süresi Maksimum 24 Saat								
		(+ ) Binary			(-) Binary			
k	%d	KârAdet	İptalAdet	Getiri	KârAdet	İptalAdet	Getiri	ToplamGetiri
<b>3</b>	<b>1.5</b>	<b>7</b>	<b>13</b>	<b>13.6%</b>	<b>1</b>	<b>2</b>	<b>1.9%</b>	<b>15.5%</b>
3	2.0	0	5	-0.2%	0	1	-0.6%	-0.8%
3	2.5	0	1	-0.2%	0	0	0.0%	-0.2%
3	3.0	0	0	0.0%	0	0	0.0%	0.0%
3	4.0	0	0	0.0%	0	0	0.0%	0.0%
<b>4</b>	<b>1.5</b>	<b>5</b>	<b>9</b>	<b>9.1%</b>	<b>0</b>	<b>0</b>	<b>0.0%</b>	<b>9.1%</b>
4	2.0	0	5	-0.2%	0	0	0.0%	-0.2%
4	2.5	0	1	-0.2%	0	0	0.0%	-0.2%
4	3.0	0	0	0.0%	0	0	0.0%	0.0%
4	4.0	0	0	0.0%	0	0	0.0%	0.0%
<b>5</b>	<b>1.5</b>	<b>5</b>	<b>4</b>	<b>9.7%</b>	<b>0</b>	<b>0</b>	<b>0.0%</b>	<b>9.7%</b>
5	2.0	0	2	-0.2%	0	0	0.0%	-0.2%
5	2.5	0	1	-0.2%	0	0	0.0%	-0.2%
5	3.0	0	0	0.0%	0	0	0.0%	0.0%
5	4.0	0	0	0.0%	0	0	0.0%	0.0%
<b>6</b>	<b>1.5</b>	<b>3</b>	<b>2</b>	<b>6.5%</b>	<b>0</b>	<b>0</b>	<b>0.0%</b>	<b>6.5%</b>
6	2.0	0	2	-0.2%	0	0	0.0%	-0.2%
6	2.5	0	1	-0.2%	0	0	0.0%	-0.2%

Saatlik (Toplam) - Bekleme Süresi Maksimum 24 Saat								
		(+ ) Binary			(-) Binary			
k	%d	KârAdet	İptalAdet	Getiri	KârAdet	İptalAdet	Getiri	ToplamGetiri
3	1.5	13	32	-23%	3	6	-10.4%	-33.2%
3	2.0	5	26	-18%	1	6	-7.9%	-26.3%
3	2.5	5	21	-12%	0	6	-9.4%	-20.9%
3	3.0	0	14	-16%	0	5	-5.3%	-20.9%
3	4.0	0	8	-3%	0	2	-2.1%	-4.7%
4	1.5	11	21	-17%	1	0	<b>1.6%</b>	-15.7%
4	2.0	7	18	-10%	0	0	0.0%	-10.0%
4	2.5	3	16	-14%	0	0	0.0%	-13.9%
4	3.0	0	12	-12%	0	0	0.0%	-12.2%
<b>4</b>	<b>4.0</b>	<b>0</b>	<b>5</b>	<b>1%</b>	<b>0</b>	<b>0</b>	<b>0.0%</b>	<b>1.2%</b>
5	1.5	8	13	-1%	0	0	0.0%	-1.0%
<b>5</b>	<b>2.0</b>	<b>6</b>	<b>12</b>	<b>10%</b>	<b>0</b>	<b>0</b>	<b>0.0%</b>	<b>9.9%</b>
5	2.5	<b>3</b>	<b>10</b>	<b>3%</b>	<b>0</b>	<b>0</b>	<b>0.0%</b>	<b>3.1%</b>
5	3.0	0	4	-2%	0	0	0.0%	-1.9%
5	4.0	0	3	-2%	0	0	0.0%	-1.5%
6	1.5	7	11	-5%	0	0	0.0%	-4.7%
6	2.0	2	11	-8%	0	0	0.0%	-7.9%
6	2.5	2	7	-2%	0	0	0.0%	-2.2%

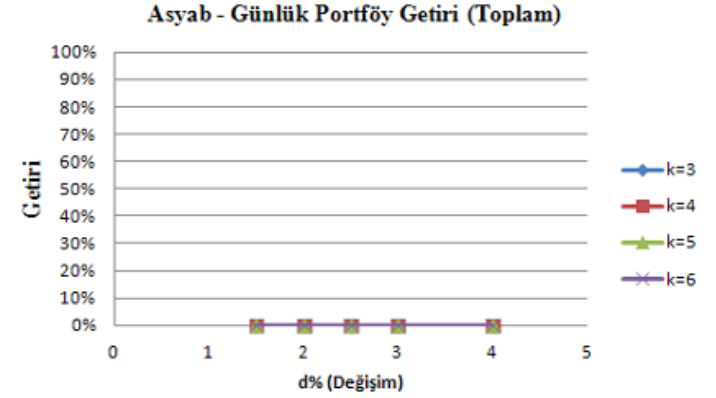
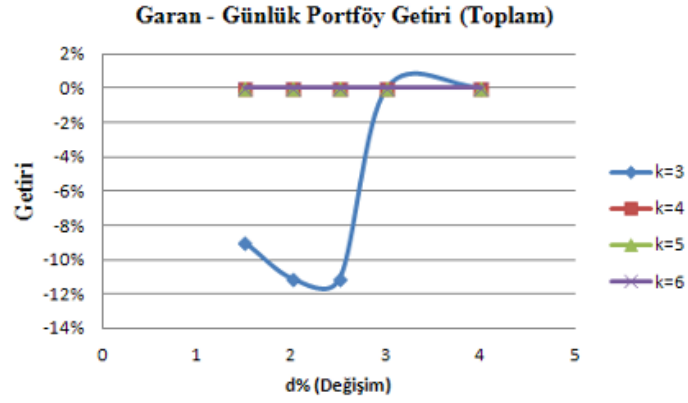
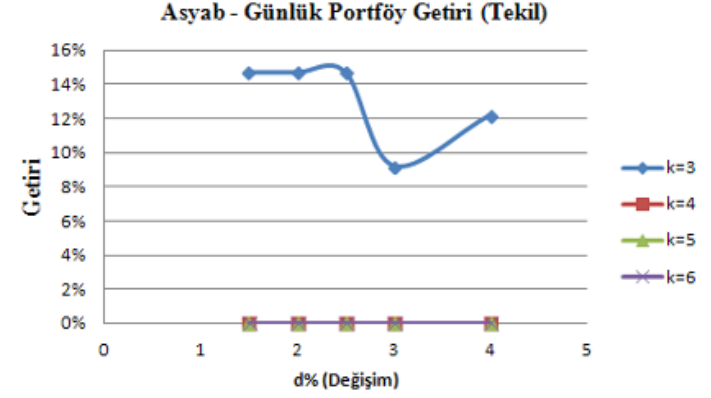
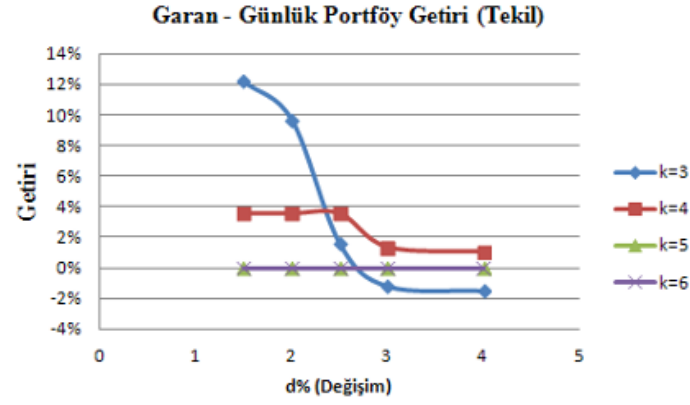
Tablo 15 - Günlük Portföy Stratejileri (Genel Haberler/Bist100)

Günlük (Tekil) - Bekleme Süresi Maksimum 7 Gün								
		(+ ) Binary			(-) Binary			
k	%d	KârAdet	İptalAdet	Getiri	KârAdet	İptalAdet	Getiri	Topl.Getiri
3	1.5	3	1	6.0%	0	1	-1.7%	4.3%
3	2.0	0	2	2.1%	0	1	0.0%	2.1%
3	2.5	0	0	0.0%	0	0	0.0%	0.0%
3	3.0	0	0	0.0%	0	0	0.0%	0.0%
3	4.0	0	0	0.0%	0	0	0.0%	0.0%
4	1.5	3	1	6.0%	0	1	-1.7%	4.3%
4	2.0	0	3	0.8%	0	1	0.0%	0.8%
4	2.5	0	1	-1.3%	0	0	0.0%	-1.3%
4	3.0	0	0	0.0%	0	0	0.0%	0.0%
4	4.0	0	0	0.0%	0	0	0.0%	0.0%
5	1.5	3	1	6.0%	0	1	-1.7%	4.3%
5	2.0	0	3	0.8%	0	1	0.0%	0.8%
5	2.5	0	1	-1.3%	0	0	0.0%	-1.3%
5	3.0	0	0	0.0%	0	0	0.0%	0.0%
5	4.0	0	0	0.0%	0	0	0.0%	0.0%
6	1.5	2	1	4.4%	0	1	-1.7%	2.7%
6	2.0	0	3	1.5%	0	1	0.0%	1.5%
6	2.5	0	2	-0.1%	0	0	0.0%	-0.1%

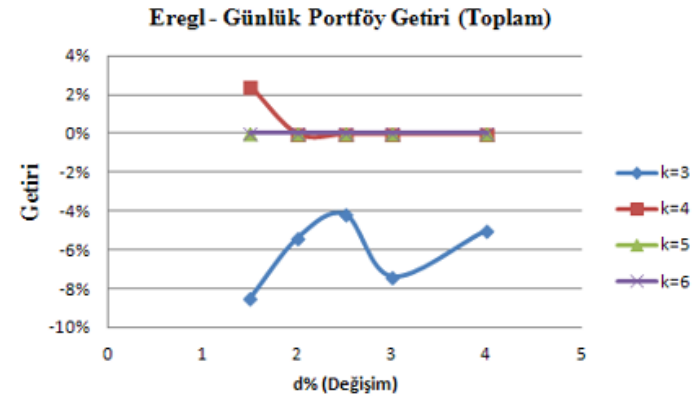
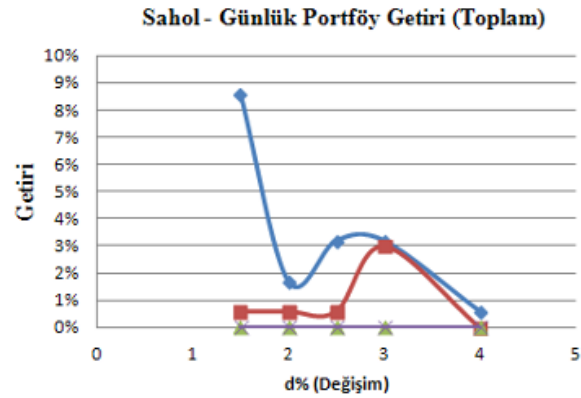
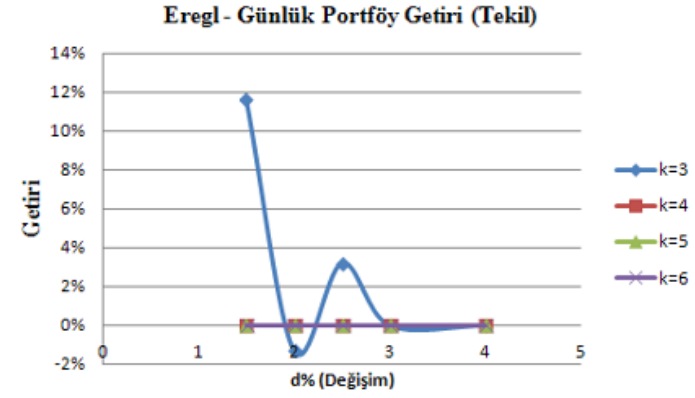
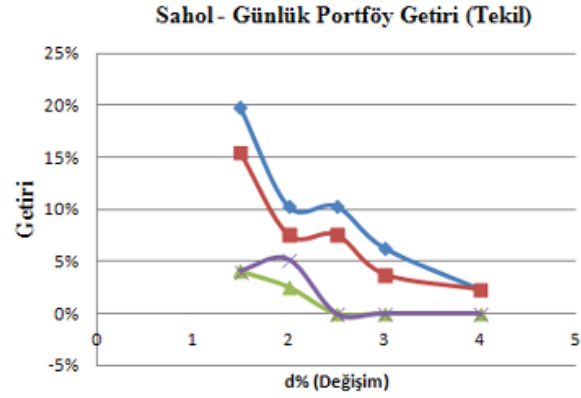
Günlük (Toplam) - Bekleme Süresi Maksimum 7 Gün								
		(+ ) Binary			(-) Binary			
k	%d	KârAdet	İptalAdet	Getiri	KârAdet	İptalAdet	Getiri	Topl.Getiri
3	1.5	1	1	-0.6%	0	1	-5.3%	-5.9%
3	2.0	1	1	1.5%	0	1	-3.7%	-2.2%
3	2.5	0	1	-3.4%	0	1	-3.7%	-7.1%
3	3.0	0	1	-3.4%	0	1	-3.7%	-7.1%
3	4.0	0	0	0.0%	0	1	-1.0%	-1.0%
4	1.5	0	3	-13.9%	0	1	-5.3%	-19.2%
4	2.0	0	3	-9.4%	0	1	-3.7%	-13.1%
4	2.5	0	1	-3.4%	0	1	-3.7%	-7.1%
4	3.0	0	1	-3.4%	0	1	-3.7%	-7.1%
4	4.0	0	0	0.0%	0	1	-3.7%	-3.7%
5	1.5	0	3	-13.9%	0	1	-5.3%	-19.2%
5	2.0	0	3	-11.8%	0	1	-3.7%	-15.5%
5	2.5	0	3	-9.4%	0	1	-3.7%	-13.1%
5	3.0	0	3	-9.1%	0	1	-3.7%	-12.8%
5	4.0	0	2	-11.6%	0	1	-3.7%	-15.3%
6	1.5	0	1	-3.4%	0	1	-5.3%	-8.7%
6	2.0	0	1	-8.4%	0	1	-3.7%	-12.1%
6	2.5	0	1	-5.8%	0	1	-3.7%	-9.5%



Şekil 19 - Bist100 Portföy Stratejisi Getiriler



Şekil 20 - Garan ve Asyab Portföy Stratejisi Getiriler



Şekil 21 - Sahol ve Eregl Portföy Stratejisi Getiriler

Şekil 19'da BIST100 Günlük Portföy Getiri (Tekil) grafiği incelendiğinde, tekil yöntemde tüm k değerleri için en yüksek getiri  $d=1.5$ , en düşük değerler ise  $d=2.5$  seviyelerinde oluşmaktadır. d değeri artırıldıkça getiriler sıfıra doğru yaklaşmaktadır. Bu durum, ilgili tarih aralığında BIST100 günlük periyot için tek seferde en büyük değişimin %3'ü geçmediğini göstermektedir.

BIST100 Günlük Portföy Getiri (Toplam) grafiğinden ise bu yöntemin doğru bir yaklaşım sergilemediği anlaşılmaktadır. Negatif getiri ağırlıklıdır.

BIST100 Saatlik Portföy Getiri (Tekil) grafiği incelendiğinde, saatlik periyotta BIST100'deki en yüksek değişimin  $d=1.5$  düzeylerinde olduğu anlaşılmaktadır. Yine toplam yönteminde günlük periyottaki durum geçerlidir.

Şekil 20 ve Şekil 21'de şirketler için hazırlanmış grafikler bulunmaktadır. Bu grafiklerde de yine toplam yöntemi genel itibari ile sonuçsuz kalmıştır. Sadece Sahol için %9 civarlarında getiri sağlamıştır (k = 3 ve k = 4 en iyi değerleri vermektedir).

Asyab için toplam yöntemi hiçbir anlam ifade etmemektedir, ancak; tekil yönteminde genel itibari ile diğer şirketlerden daha çok getiri elde etmiştir. Dolayısıyla ani değişimlerin bu hisse senedi fiyatları için sıklıkla yaşandığı söylenebilir.

Eregl ve Garan, toplam yöntemi ile çoğunlukla negatif getiri elde etmektedir. Garan, tekil yöntemde daha dengeli görünmektedir. Eregl ise en iyi getiriyi k = 3,  $d=1.5$  değerleriyle sağlamıştır. Diğer değerler için ani sıçramalar görünmemektedir.

Sahol için tekil grafik incelendiğinde, genel itibari ile pozitif getiri elde ettiği ve d değeri artırıldıkça getirinin azaldığı izlenmektedir. Sahol'ün her iki grafiğinde de negatif bir eğilim gözlemlenmemiştir.

Tüm grafiklerden anlaşılıyor ki, %1.5'lik bir değişim her bir yatırım aracında, tek seferde oluşabilecek bir değer ve en iyi getiriyi sağlayan parametredir.

Tablo 14 ve Tablo 15'de, Portföy-I stratejisi için 2014.01.01 - 2014.08.15 tarih aralığındaki eğitim sonuçları, Tablo 16'da ise 2014.09.01 - 2014.12.31 tarihleri için test sonuçları yer almaktadır (sadece Günlük-Tekil tablosu gösterilmiştir).

Burada eğitim tarih aralığı ile benzer veriler elde edildiği görülmektedir.

Tablo 16 - Portföy-I Test Tarih Aralığı Günlük (Tekil) Getiriler

Günlük (Tekil) - Bekleme Süresi Maksimum 7 Gün								
k	%d	(+ ) Binary			(- ) Binary			ToplamGetiri
		KarAdet	İptalAdet	Getiri	KarAdet	İptalAdet	Getiri	
3	1.5	1	2	2.2%	0	0	0.0%	2.2%
3	2.0	1	1	2.6%	0	0	0.0%	2.6%
3	2.5	0	1	0.4%	0	0	0.0%	0.4%
3	3.0	0	0	0.0%	0	0	0.0%	0.0%
3	4.0	0	0	0.0%	0	0	0.0%	0.0%
4	1.5	0	2	0.0%	0	0	0.0%	0.0%
4	2.0	0	1	0.4%	0	0	0.0%	0.4%
4	2.5	0	1	0.4%	0	0	0.0%	0.4%
4	3.0	0	0	0.0%	0	0	0.0%	0.0%
4	4.0	0	0	0.0%	0	0	0.0%	0.0%
5	1.5	1	1	3.5%	0	0	0.0%	3.5%
5	2.0	0	1	0.4%	0	0	0.0%	0.4%
5	2.5	0	1	0.4%	0	0	0.0%	0.4%
5	3.0	0	0	0.0%	0	0	0.0%	0.0%
5	4.0	0	0	0.0%	0	0	0.0%	0.0%
6	1.5	1	1	3.5%	0	0	0.0%	3.5%
6	2.0	0	1	0.4%	0	0	0.0%	0.4%
6	2.5	0	1	0.4%	0	0	0.0%	0.4%
6	3.0	0	0	0.0%	0	0	0.0%	0.0%
6	4.0	0	0	0.0%	0	0	0.0%	0.0%

## Portföy Stratejisi II

Portföy Stratejisi I'in sonuçlarına bakıldığı zaman ani sıçramaları doğru tespit ettiği, fakat getiri konusunda yetersiz kaldığı görülmüştür. Yine aynı hipotez üzerinden yola çıkılarak, getiri odaklı daha basit bir strateji geliştirilmiştir.

Portföy-II için, Portföy-I'deki tüm parametreler ve yöntemler kullanılmıştır. Bekleme süresi olarak çalışma verisindeki kayıt sayısı temel alınmıştır. Liste sonuna kadar gelindiğinde portföy elden çıkarılmıyorsa iptal olarak varsayılmıştır. Portföy-I için bekleme süresi kısıtlı tutulmuştur.

**Portföy-II çalışma prensibi:** Günlük veya saatlik veriler üzerinde, giriş parametrelerinde belirtilen tüm elemanlar için: k adet tekrar eden desen yakalandığında al/sat (yeni portföy oluşturma), %d değerine ulaşıldığında al/sat işlemi (portföyü elden çıkartma) yapılır. Eğer bekleme süresi dolduğunda (normal



şartlarda kısa vade iptal işlemi bulunmamaktadır) %d değerine ulaşamaz ise kâr veya zarar durumuna bakılmaksızın portföy elden çıkarılır. Şekil 22’de akış şeması gösterilmiştir (Portföy II - sağ).

Tablo 17’de Portföy-II için saatlik test sonuçları listelenmiştir:

Tekil yöntemde  $k = 3$ ,  $d = 2$  ile %153 getiri sağlamıştır. Buradan anlaşılıyor ki test aralığında tek seferde yakaladığı en yüksek fiyat değişimi %2 civarlarındadır. Hem pozitif ve hem de negatif binary kazanç sağlamıştır.

Toplam yönteminde ise negatif binary’de;  $k = 3$ ,  $d = 4$  için %31.6’lık getiri sağlamış ve hiç bir adımda zarar etmemiştir. Toplam yönteminde pozitif binary kullanılmaması gerektiği anlaşılmıştır.

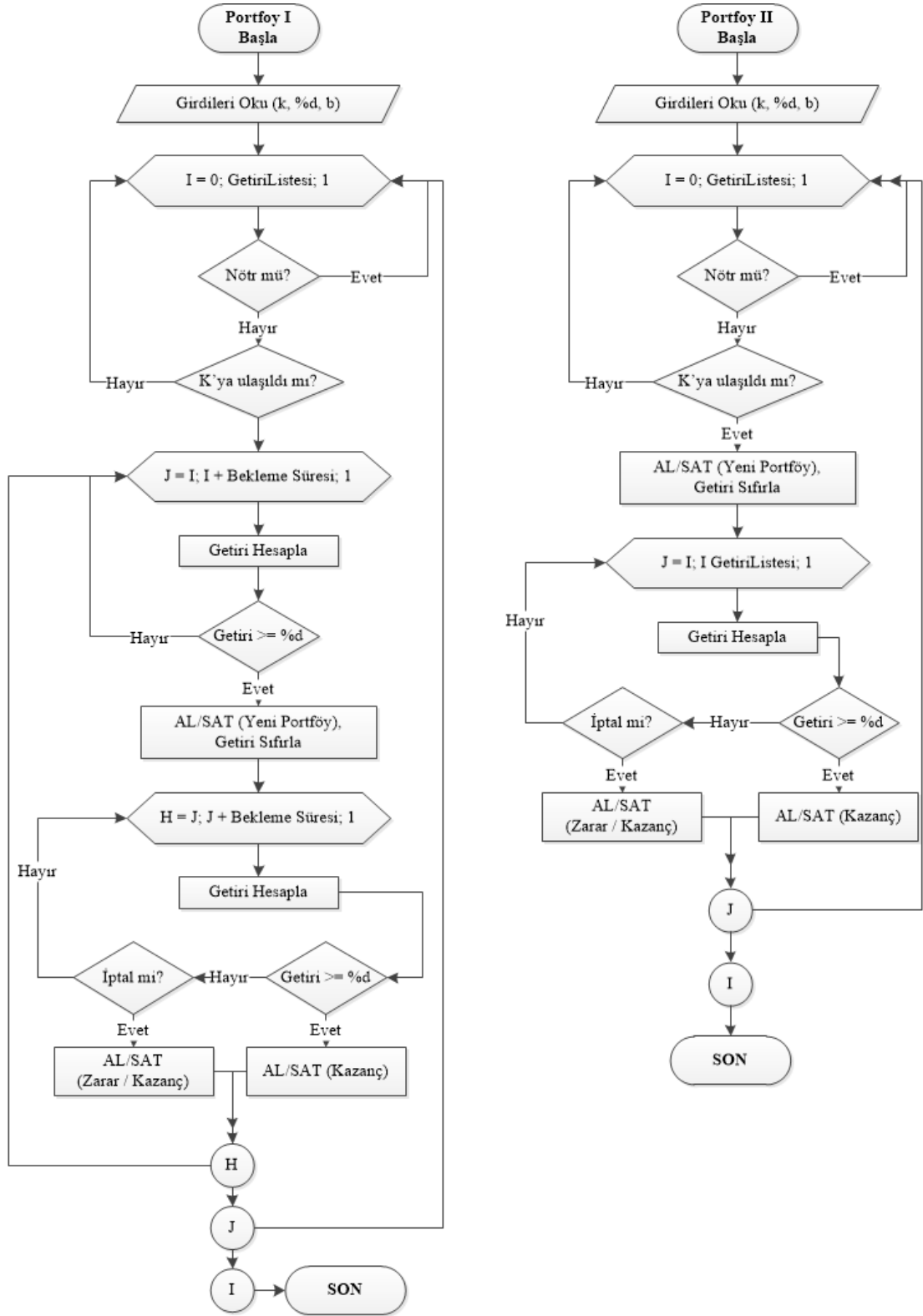
Günlük getiriler incelendiğinde, en yüksek %16’lık bir kazanç söz konusudur.

Portföy Stratejisi II ile eğitim tarih aralığı için (2014.01.01 - 2014.08.15) çalışma yapıldığında, yine saatlik tekil yöntemde en yüksek %248’lik getiri elde edilmiştir ( $k=3$ ,  $d=2$ ).

Şekil 23’de, Portföy Stratejisi I ve II için eğitim tarih aralığında, saatlik getiri karşılaştırması yapılmıştır. Portföy I tekil yöntem için  $d=2$ , Portföy II tekil yöntem için  $d=3$  en yüksek saatlik değişimi yakaladığı değerleri göstermektedir.

Toplam yöntemde, Portföy I için %d arttıkça, negatif getiri sifira yaklaşırken, Portföy II için sıfırdan negatif yönde uzaklaşmaktadır.

Bu çalışma sonucunda, haberler ve getiri arasında doğrusal bir ilişki çıkmazken, portföy uygulaması ile haberlerin eğilimini eşleştirerek başarılı bir getiri elde edilebileceği anlaşılmaktadır. Uygulamada (+) ve (-) binary’ler için tepki veya normal seyir alım/satım işlemlerine göre ayrıştırma yapılırsa getiri yüzdesi daha da artacaktır.

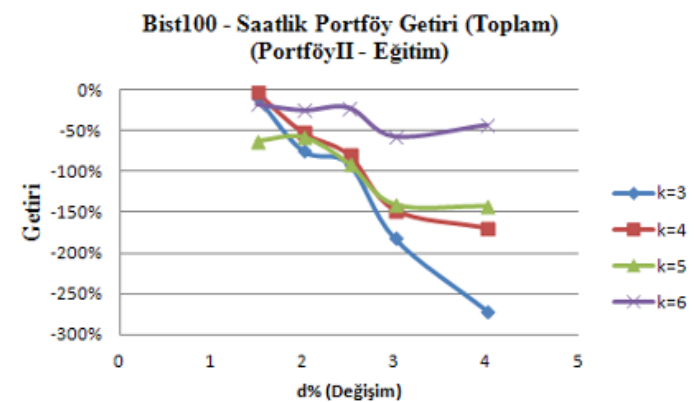
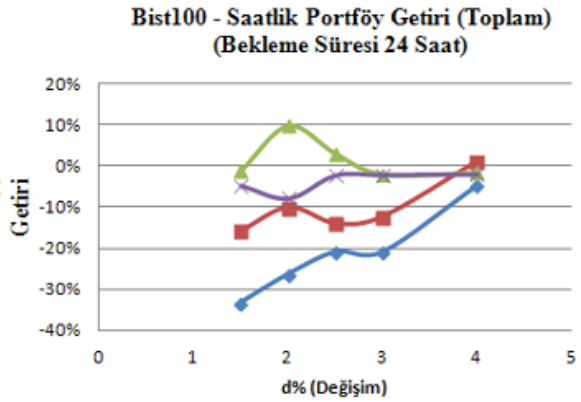
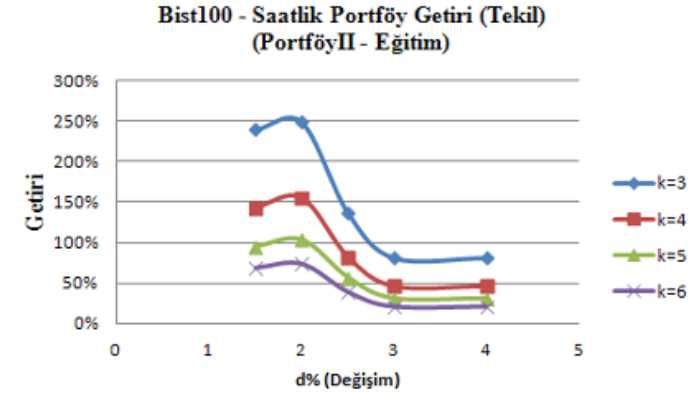
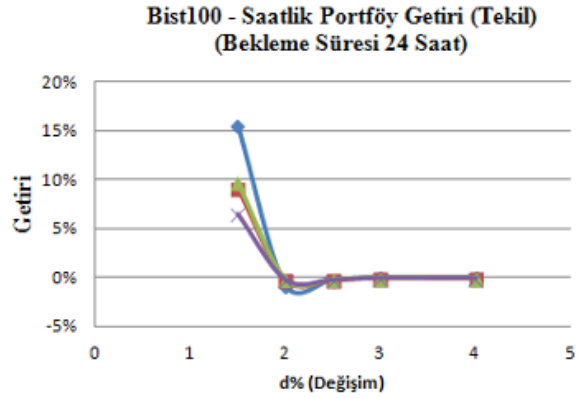


Şekil 22 - Portföy Stratejisi I ve II - Akış Şeması

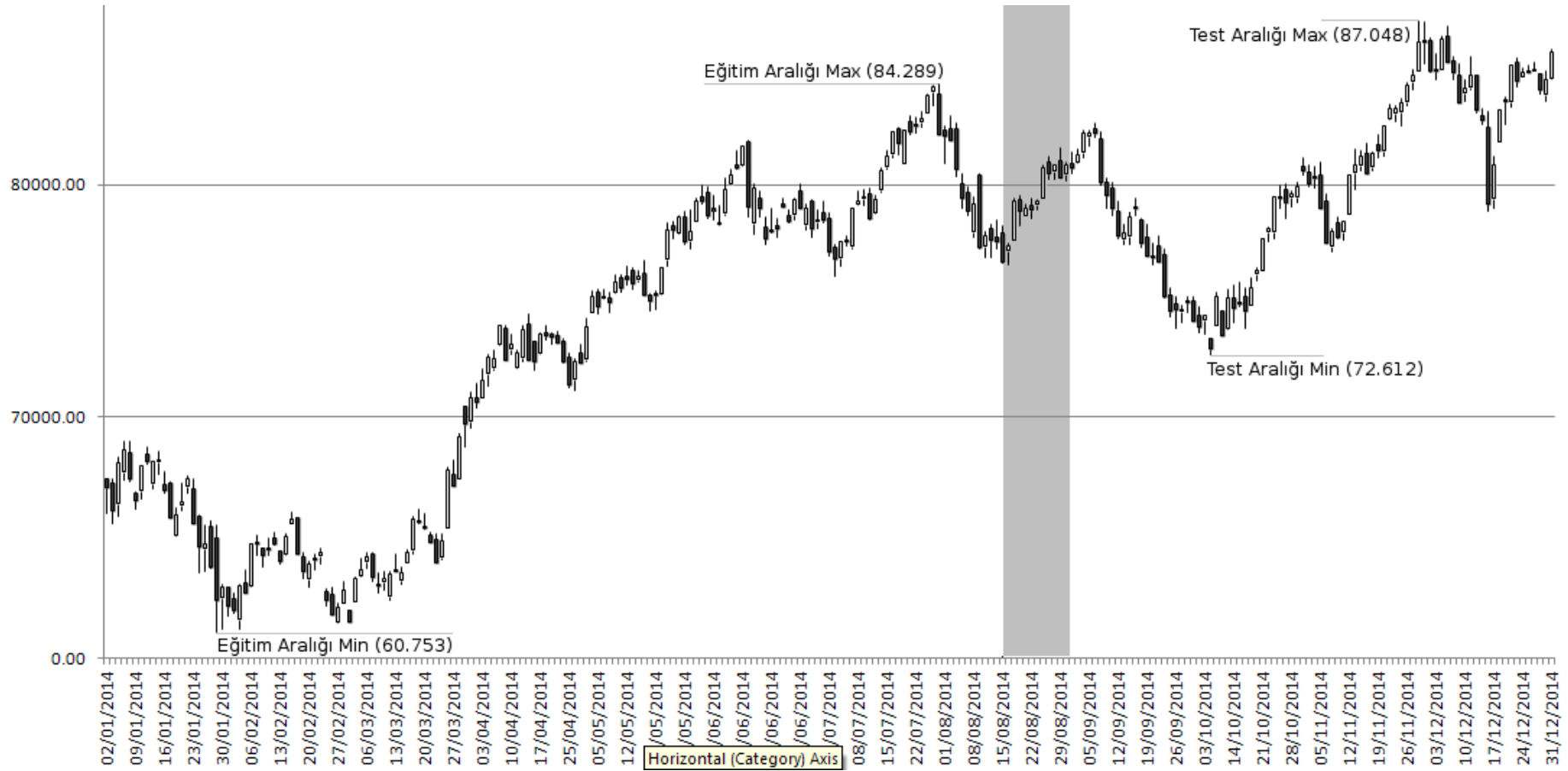
Tablo 17 - Portföy II Test Tarih Aralığı Saatlik Getiriler

Saatlik (Tekil) - Test Tarih Aralığı								
		(+ ) Binary			(- ) Binary			
k	%d	KârAdet	İptalAdet	Getiri	KârAdet	İptalAdet	Getiri	Topl.Getiri
<b>3</b>	<b>1.5</b>	<b>65</b>	<b>7</b>	<b>114.0%</b>	<b>8</b>	<b>1</b>	<b>15.2%</b>	<b>129.3%</b>
<b>3</b>	<b>2.0</b>	<b>65</b>	<b>7</b>	<b>136.7%</b>	<b>8</b>	<b>1</b>	<b>16.7%</b>	<b>153.4%</b>
3	2.5	0	72	-64.8%	0	9	-8.1%	-72.9%
3	3.0	0	72	-64.8%	0	9	-8.1%	-72.9%
3	4.0	0	72	-64.8%	0	9	-8.1%	-72.9%
<b>4</b>	<b>1.5</b>	<b>47</b>	<b>6</b>	<b>80.2%</b>	<b>3</b>	<b>0</b>	<b>6.0%</b>	<b>86.2%</b>
<b>4</b>	<b>2.0</b>	<b>47</b>	<b>6</b>	<b>98.0%</b>	<b>3</b>	<b>0</b>	<b>6.6%</b>	<b>104.6%</b>
4	2.5	0	53	-47.7%	0	3	-2.7%	-50.4%
4	3.0	0	53	-47.7%	0	3	-2.7%	-50.4%
4	4.0	0	53	-47.7%	0	3	-2.7%	-50.4%
<b>5</b>	<b>1.5</b>	<b>30</b>	<b>5</b>	<b>50.9%</b>	<b>1</b>	<b>0</b>	<b>1.9%</b>	<b>52.8%</b>
<b>5</b>	<b>2.0</b>	<b>30</b>	<b>5</b>	<b>61.5%</b>	<b>1</b>	<b>0</b>	<b>2.2%</b>	<b>63.7%</b>
5	2.5	0	35	-31.5%	0	1	-0.9%	-32.4%
5	3.0	0	35	-31.5%	0	1	-0.9%	-32.4%
5	4.0	0	35	-31.5%	0	1	-0.9%	-32.4%
<b>6</b>	<b>1.5</b>	<b>21</b>	<b>5</b>	<b>35.4%</b>	<b>1</b>	<b>0</b>	<b>1.9%</b>	<b>37.3%</b>
<b>6</b>	<b>2.0</b>	<b>21</b>	<b>5</b>	<b>41.7%</b>	<b>1</b>	<b>0</b>	<b>2.2%</b>	<b>43.9%</b>
6	2.5	0	26	-23.4%	0	1	-0.9%	-24.3%

Saatlik (Toplam) - Test Tarih Aralığı								
		(+ ) Binary			(- ) Binary			
k	%d	KârAdet	İptalAdet	Getiri	KârAdet	İptalAdet	Getiri	Topl.Getiri
3	1.5	54	18	-54.2%	8	1	11.4%	-42.8%
3	2.0	50	22	-69.0%	8	1	18.1%	-50.9%
3	2.5	46	26	-94.0%	8	1	20.5%	-73.5%
3	3.0	41	31	-132.5%	8	1	23.7%	-108.8%
3	4.0	34	38	-161.1%	8	1	31.6%	-129.5%
4	1.5	36	17	-72.3%	3	0	4.8%	-67.5%
4	2.0	32	21	-115.8%	3	0	7.6%	-108.2%
4	2.5	31	22	-112.3%	3	0	8.6%	-103.7%
4	3.0	26	27	-159.5%	3	0	10.0%	-149.5%
4	4.0	21	32	-178.7%	3	0	12.4%	-166.3%
5	1.5	24	11	-44.5%	1	0	1.5%	-43.0%
5	2.0	21	14	-62.4%	1	0	2.1%	-60.3%
5	2.5	19	16	-80.1%	1	0	2.6%	-77.5%
5	3.0	16	19	-95.3%	1	0	3.4%	-91.9%
5	4.0	13	22	-119.1%	1	0	4.5%	-114.6%
6	1.5	17	9	-27.5%	1	0	1.7%	-25.8%
6	2.0	14	12	-46.8%	1	0	2.0%	-44.8%
6	2.5	12	14	-66.0%	1	0	2.5%	-63.5%



Şekil 23 - Portföy I - Portföy II Saatlik Eğitim Getiri Karşılaştırma



Şekil 24 - Bist100 Günlük Fiyat Değişim

### 3.4. BIST100 Analiz

Tablo 18’de BIST100 için genel istatistikler listelenmiştir. Portföy uygulamamızda alınan sonuçlar bu tablo ile karşılaştırıldığında saatlik periyot için BIST100’ün yaptığı maksimum değişim yakalanmıştır. Ancak, günlük getiriler için bunu söyleyemeyiz. %5 ve -%4’lük değişimler günlük hesaplamalarda haber eğilimleri ile eşleştirilememiştir.

Yatırımcı, günlük periyotta BIST100’e 02.01.2014 tarihinde girip, 15.08.2014 tarihinde çıkmış olsaydı elde edeceği getiri %13.8 olacaktı. Portföy-I günlük stratejimiz ile %4.3 kazanmıştır, ancak saatlik stratejimiz ile %15.5 kazanmıştır. Yani Portföy-I Saatlik (Tekil) BIST100’den iyi bir getiri elde etmiştir.

Yatırımcı, günlük periyotta BIST100’e 01.09.2014 tarihinde girip 31.12.2014 tarihinde çıkmış olsaydı elde edeceği getiri %6.5 olacaktı. Portföy-I günlük stratejimiz ile %3.5 kazanmıştır, ancak saatlik stratejimiz ile %2 kazanmıştır. Bu durumda test aralığında BIST100 kazandırmış olacaktı.

Portföy-II stratejimiz ile eğitim aralığında günlük (tekil) %35.2, saatlik (tekil) %248 getiri sağlayarak BIST100’ün oldukça üzerinde kazandıracaktır. Yine test aralığında günlükte %16.2, saatlikte %153 getiri ile BIST100’den fazla kazandıracaktır.

Tablo 18 - Bist100 Genel Veriler

<b>BIST100</b>	<b>Öğrenme Aralığı</b>	<b>Test Aralığı</b>
Başlangıç Tarihi	02.01.2014	01.09.2014
Bitiş Tarihi	15.08.2014	31.12.2014
Başlangıç Değeri	67.367	80.455
Son Değer	76.692	85.721
Minimum	60.753	72.612
Maksimum	84.289	87.048
Saatlik Maksimum Artış	2.38%	2.21%
Saatlik Maksimum Düşüş	-2.58%	-2.23%
Günlük Maksimum Artış	5%	3%
Günlük Maksimum Düşüş	-3%	-4%
Aylık Maksimum Artış	11.48%	7.52%
Aylık Maksimum Düşüş	-6.70%	-5.20%

### 3.5. Kelime Analiz

Tablo 19'da eğitim ve test tarih aralığındaki, ham ve işlenmiş haberlerin içerisinde bulunan toplam kelime adetleri gösterilmiştir. Tablo 20'de ise Türkiye Ekonomik Psikoloji Sözlüğünde tanımlanmış kelimelerin, tarama sonuçları gösterilmiştir. Sonuçlar içerisinde Genel ve Şirket kategorilerinin tümü yer almaktadır.

Tablo 19 - Ham ve İşlenmiş Haberlerin İçeriğindeki Kelime Sayıları

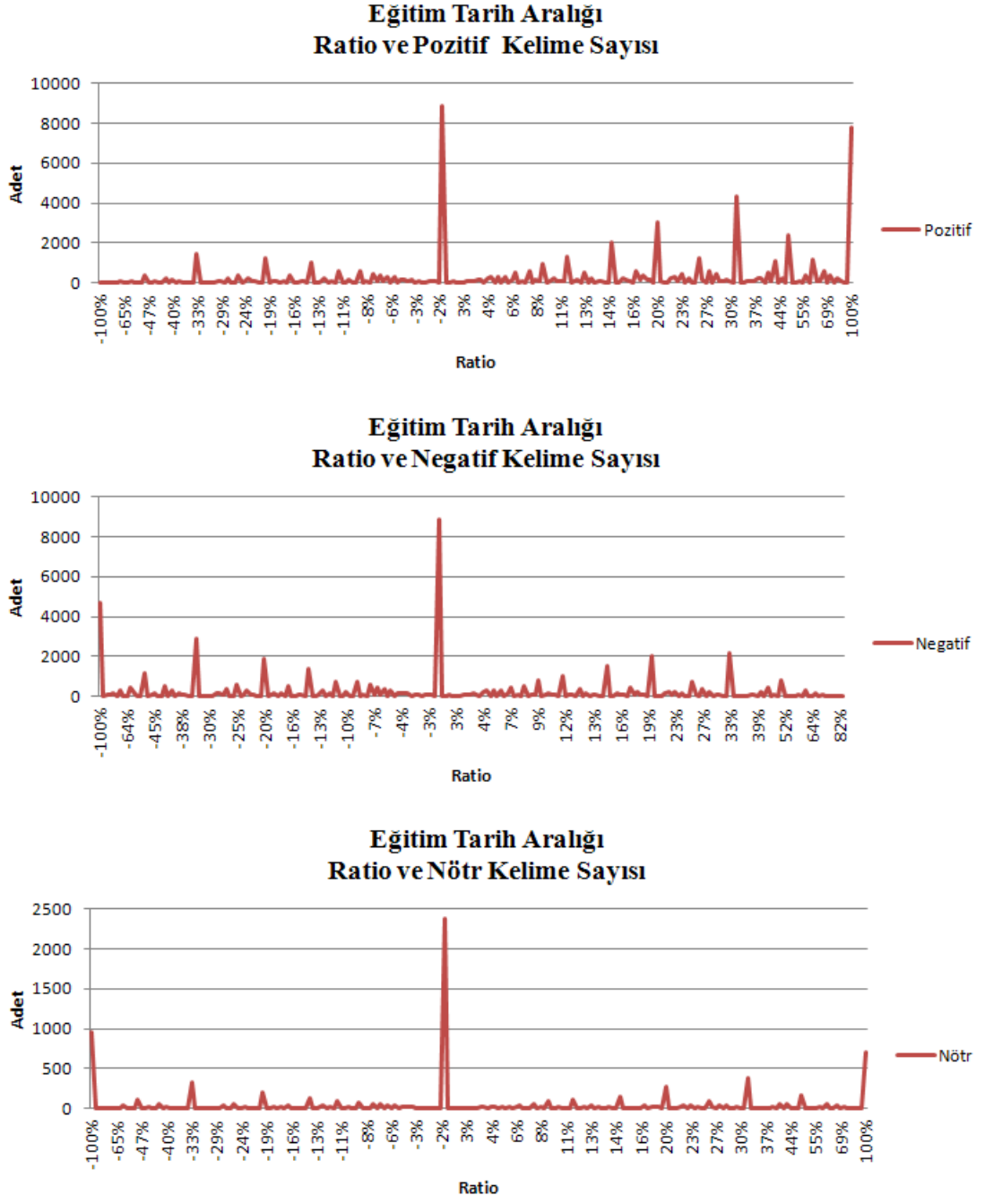
	Ham Veri			İşlenmiş Veri		
	Toplam Haber Sayısı	Toplam Kelime Sayısı	Haber Başına Düşen Kelime Sayısı	Toplam Haber Sayısı	Toplam Kelime Sayısı	Haber Başına Düşen Kelime Sayısı
<b>Eğitim</b>	34.741	12.148.407	350	21.244	8.090.935	380
<b>Test</b>	17.137	5.874.144	343	11.106	4.673.015	420

Tablo 20 - Haber Kelime İstatistikleri

	Eğitim Tarih Aralığı	Test Tarih Aralığı
Toplam Yakalanan Kelime Sayısı	120.556	64.165
Haber Başına Yakalanan Ortalama Kelime Sayısı	5.7	5.7
Toplam Yakalanan Pozitif Kelime Sayısı	60.507	32.302
Toplam Yakalanan Negatif Kelime Sayısı	51.681	26.817
Toplam Yakalanan Nötr Kelime Sayısı	8.368	5.046
Bir Haberde Yakalanan En Fazla Kelime Sayısı	67	78
Bir Haberde Yakalanan En Az Kelime Sayısı	1	1
Bir Haberde Yakalanan En Fazla Pozitif Kelime Sayısı	34	38
Bir Haberde Yakalanan En Fazla Negatif Kelime Sayısı	32	36
Bir Haberde Yakalanan En Fazla Nötr Kelime Sayısı	5	8

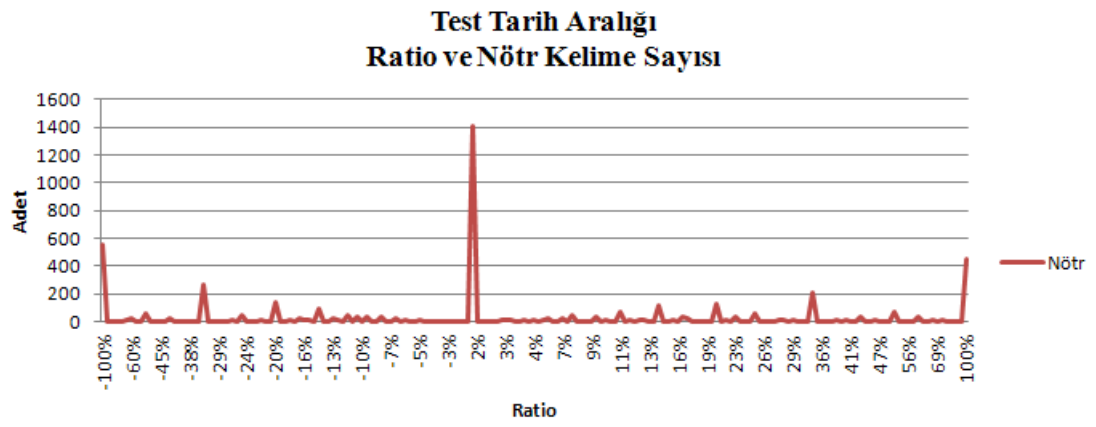
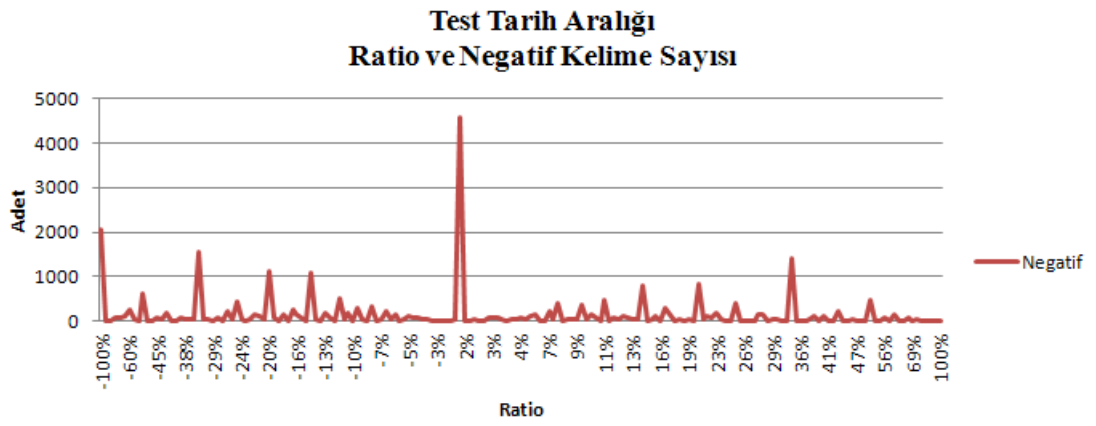
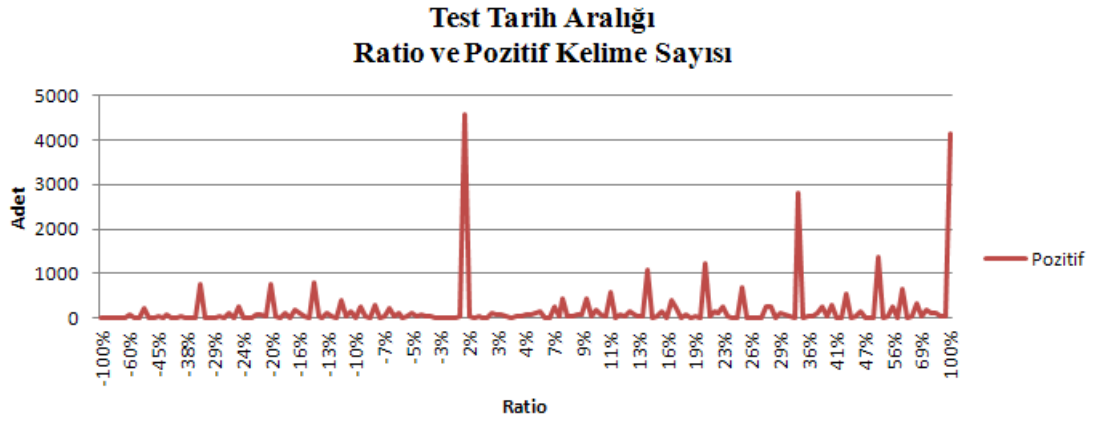
Tablo 20 incelendiğinde, haber başına düşen yakalanan kelime oranlarında, test tarih aralığı için bir artış görülmektedir. Sözlük tanımlama aşamasında eğitim tarih aralığında bulunan haberlerden yararlanılmış olmasına rağmen ileri tarihli haberler için de bu tanımlamaların geçerli olduğu anlaşılmaktadır.

Haber başına yakalanan ortalama kelime sayısının artırılması için sözlüğün geliştirilmesi gerekmektedir.



Şekil 25 - Eğitim Tarih Aralığı Ratio ve Kelime Sayıları





Şekil 26 - Test Tarih Aralığı Ratio ve Kelime Sayıları

Şekil 25 ve Şekil 26'ya göre, yakalanan kelime sayısı arttıkça  $-3\%$  ile  $+3\%$  Ratio oranlarına yaklaşılmakta ve haberlerin eğilimi konusunda oldukça hassas bir aralığa gelinmektedir.

#### 4. Sonuç

Türkiye’de geriye yönelik veri arşivlerinin elde edilmesi ile ilgili genel olarak zorluklar yaşanmaktadır. Çalışmamız için gerekli haber arşivlerinin temini konusunda, başvurulmuş birçok kurumdan ret cevabı alınmıştır. Kurumlar verileri paylaşmamış veya fiziksel koşullar nedeniyle uzun süreli arşiv tutmadıklarını öne sürmüşlerdir. Gerekli veriler yazılan bir uygulama yardımı ile internet üzerinden elde edilmiş, bu da yürütülen çalışmanın süresinin oldukça uzamasına neden olmuştur.

Var olan modellere bağımlı kalınmadan, yenilikçi yaklaşımlarla farklı modeller geliştirilmiş ve literatürde bulunmayan, özgün çalışmalar yapılmıştır. Bu bağlamda Bag-of-Words modeli geliştirilerek sorgulama özelliği eklenmiştir. Böylelikle bir metin üzerinde çalışılırken öğrenen sistemlere gerek kalmaksızın, bu sistemler kadar başarılı sonuçlar elde edilmiştir.

Yeni geliştirilen model ile birlikte, Türkiye için henüz rastlanmamış olan EkonomiK Psikoloji Sözlüğünün temeli atılmıştır. Doğal dil işleme ile ilgili çalışmalar mevcuttur, ancak ekonomi alanında böyle bir çalışma bulunmamaktadır. Bu sözlük geliştirilmeye açık durumdadır. Sözlük ile ilgili çalışmalarda, alanında uzmanların görüşleri alınabilir.

Sözlük tanımlanırken yapılan çalışmada varılan genel kanı ve istatistiklerimiz gösteriyor ki, psikolojik olarak negatif eğilimli tanımlamalar yapılmaktadır. Sözlükte tanımlanmış olan kelime gruplarına; genel kategorisi için %56 oranında negatif, %39 pozitif giriş yapıldığı gözlemlenmiştir. Buna rağmen işlenen haberlerin eğilimine bakıldığında, bunun tam tersi bir görüntü oluşmuş ve pozitif ağırlıklı bir veri elde edilmiştir. Bu da gösteriyor ki, incelediğimiz tarih aralığında piyasada pozitif bir algı oluşturulmak istenmektedir.

Günlük ve saatlik periyotta; Türkiye ekonomi haberlerinin, pozitif veya negatif olarak ayrıştırılarak, BIST100 ve şirket hisselerinin fiyat değişimleri üzerinde etkisi olduğuna dair ampirik bir kanıt bulunamamıştır.

Literatürde kullanılan korelasyon, regresyon vb. analiz yöntemleri ile, haber eğilimleri ve fiyat değişimleri arasında bir bağlantı kurulamamıştır. Daha karmaşık bir model kurularak iki farklı Portföy Stratejisi uygulaması geliştirilmiştir. İkinci strateji, yüksek getiri konusunda umut veren sonuçlar üretmiştir. Portföy stratejisi

uygulaması daha da geliştirilerek, minimum portföyle risksiz getiri elde etme konusunda çalışmalar yürütülebilir.

Sadece Türkiye ekonomi haberleri ile çalışmak genel anlamda istenen sonucu vermemiştir. İleriki çalışmalarda, siyasi haberlerin ve dünya olayları kategorilerinin de veri setine dâhil edilmesi düşünülmektedir. Haberlerin daha kısa periyotlarla yorumlanmasının anlık fiyat hareketlerine etkisi incelenebilir. Anlık haber akışları ile ilgili olarak yine arşiv sıkıntısı nedeniyle, çalışmalar, geçmiş veri ile değil anlık yapılabilir. Sonuçta geliştirilmiş olan modelimiz öğrenme süreci gerektirmemektedir, anlık veri üzerinden de çalışabilmektedir.

## Kaynakça

- Amihud, Y., & Wohl, A. (2003). Political News And Stock Prices: The Case Of Saddam Hussein Contracts. *Journal of Banking & Finance* , 1185–1200.
- Baray, A. (2003). Entropi ve Karar Verme. *Yönetim*, 44 , 7-21.
- Blasco, N., Corredor, P., Rio, C., & Santamaria, R. (2005). Bad News And Dow Jones Make The Spanish Stocks Go Around. *European Journal of Operational Research* 163 , 253-275.
- Boyd, J. H., Hu, J., & Jagannathan, R. (2005). The Stock Market's Reaction to Unemployment News: Why Bad News Is Usually Good for Stocks. *The Journal Of Finance* , 649–672.
- Budak, A., & Fıkırkoca, E. (2011). *Yatırım Yaparken Sorularla Türkiye Sermaye Piyasası*. İstanbul: Paragraf Basım Sanayi A.Ş.
- Carvalho, C., Klagge, N., & Moench, E. (2011). The Persistent Effects Of A False News Shock. *Journal of Empirical Finance* , 597–615.
- Chan, S., & Franklin, J. (2011). A Text-Based Decision Support System For Financial Sequence Prediction. *Decision Support Systems* , 189–198.
- Chen, W., & Szeto, K. (2012). Mixed Time Scale Strategy In Portfolio Management. *International Review of Financial Analysis - Volume 23* , 35–40.
- Fan, J., & Su, K. (1994). An Efficient Algorithm For Matching Multiple Patterns. J. Aoe içinde, *Computer Algorithms: String Pattern Matching Strategies* (s. 91-103). John Wiley & Sons.
- Fortuny, E., Smedt, T., Martens, D., & Daelemans, W. (2014). Evaluating And Understanding Text-Based Stock Price Prediction Models. *Information Processing And Management* , 426–441.
- Fung, G., Yu, j., & Lam, W. (2002). News Sensitive Stock Trend Prediction. *Advances in Knowledge Discovery and Data Mining* (s. 481-493). içinde
- Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance* 18 , 321-340.
- Harris, Z. (1954). Distributional Structure. *Word* 10 (2/3) , 146-162.
- Hasib, S., Motwani, M., & Saxena, A. (2013). Importance of Aho-Corasick String Matching Algorithm in Real World Applications. *International Journal of Computer Science and Information Technologies* , 467-469.
- Hurwitz, R. (2002). *Harvard University*. Descriptions of Inquirer Categories and Use of Inquirer Dictionaries: <http://www.wjh.harvard.edu/~inquirer/homecat.htm> adresinden alınmıştır
- Jaynes, E. (1957). Information Theory And Statistical Mechanics. *Physical Review*, 106 , 620-630.
- Joulin, A., Lefevre, A., Grunberg, D., & Bouchaud, J. (2008). Stock Price Jumps: News and Volume Play a Minor Role. *Wilmott Magazine* , 1-7.

- Kumar, A., & Chaudhary, A. (2004). Correlation And Regression. A. Kumar, & A. Chaudhary içinde, *Kirshna's Descriptive Statistics: (Statistical Methods)* (s. 187-232). Krishna Prakashan Media.
- Li, Q., Wang, T., Li, P., Liu, L., & Gong, Q. (2014). The Effect Of News And Public Mood On Stock Movements. *Information Sciences* , 826–840.
- Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News Impact On Stock Price Return Via Sentiment Analysis. *Knowledge-Based Systems* .
- McDonald, B. (2013). *Textual Analysis*. University of Notre Dame: [http://www3.nd.edu/~mcdonald/Word\\_Lists.html](http://www3.nd.edu/~mcdonald/Word_Lists.html) adresinden alınmıştır
- McQuail, D. (1987). *Mass Communication Theory*. Thousand Oaks, CA, US: Sage Publications.
- Microsoft. (2014). *Developer Tools*. Microsoft Download Center: <http://www.microsoft.com/tr-tr/download/developer-tools.aspx> adresinden alınmıştır
- Microsoft. (2014). *LINQ (Language-Integrated Query)*. Microsoft Developer Network: <http://msdn.microsoft.com/en-us/library/bb397926.aspx> adresinden alınmıştır
- Microsoft. (2014). *MSHTML*. MSDN: <http://msdn.microsoft.com/en-us/library/bb508515%28v=vs.85%29.aspx> adresinden alınmıştır
- MSDN. (2014). *Language-Integrated Query*. Microsoft Developer Network: <http://msdn.microsoft.com/en-us/library/bb397926.aspx> adresinden alınmıştır
- Nassirtoussi, A., Aghabozorgi, S., Wah, T., & Ngo, D. (2014). Text Mining For Market Prediction: A Systematic Review. *Expert Systems With Applications* , 7653–7670.
- Nguyen, V., & Claus, E. (2013). Good News, Bad News, Consumer Sentiment And Consumption. *Journal of Economic Psychology* 39 , 426–438.
- Northedge, R. (2006). *Code Project*. Maximum Entropy Modeling Using SharpEntropy: <http://www.codeproject.com/Articles/11090/Maximum-Entropy-Modeling-Using-SharpEntropy> adresinden alınmıştır
- Obiwan, D. (2012). *Html Agility Pack*. Codeplex: <http://htmlagilitypack.codeplex.com/> adresinden alınmıştır
- Oracle. (2014). *MySQL*. MySQL: <http://www.mysql.com> adresinden alınmıştır
- Petricek, T. (2005). *Aho-Corasick string matching in C#*. Code Project: <http://www.codeproject.com/Articles/12383/Aho-Corasick-string-matching-in-C> adresinden alınmıştır
- Piper, P. (2006). *ObjectListView*. sourceforge.net: <http://objectlistview.sourceforge.net/cs/index.html> adresinden alınmıştır
- Rigel, N. (2000). *İleti Tasarımında Haber*. İstanbul: Der Yayınları.
- Schumaker, R., & Chen, H. (2009). A Quantitative Stock Prediction System Based On Financial News. *Information Processing and Management* 45 , 571-583.
- Schumaker, R., Zhang, Y., Huang, C., & Chen, H. (2012). Evaluating Sentiment In Financial News Articles. *Decision Support Systems* , 458–464.

Veronesi, P. (1999). Stock Market Overreactions To Bad News In Good Times: A Rational Expectations Equilibrium Model. *Review of Financial Studies* 12.5 , 975-1007.

Wikipedia. (2014). *Bag-of-words model*. Wikipedia:  
[http://en.wikipedia.org/wiki/Bag-of-words\\_model](http://en.wikipedia.org/wiki/Bag-of-words_model) adresinden alınmıştır

Yu, L., Wu, J., Chang, P., & Chu, H. (2013). Using A Contextual Entropy Model To Expand Emotion Words And Their Intensity For The Sentiment Classification Of Stock Market News. *Knowledge-Based Systems* , 89–97.